VISIT…

LANZAROTE Caliente.COM

edited by DANIEL J. LEVITIN

FOUNDATIONS OF COGNITIVE PSYCHOLOGY

CORE *readings*

Preface

Daniel J. Levitin

What Is Cognition?

Cognition encompasses the scientific study of the human mind and how it processes information; it focuses on one of the most difficult of all mysteries that humans have addressed. The mind is an enormously complex system holding a unique position in science: by necessity, we must use the mind to study itself, and so the focus of study and the instrument used for study are recursively linked. The sheer tenacity of human curiosity has in our own lifetimes brought answers to many of the most challenging scientific questions we have had the ambition to ask. Although many mysteries remain, at the dawn of the twenty-first century, we find that we do understand much about the fundamental laws of chemistry, biology, and physics; the structure of space-time, the origins of the universe. We have plausible theories about the origins and nature of life and have mapped the entire human genome. We can now turn our attention inward, to exploring the nature of thought, and how our mental life comes to be what it is.

There are scientists from nearly every field engaged in this pursuit. Physicists try to understand how physical matter can give rise to that ineffable state we call consciousness, and the decidedly nonphysical "mind stuff" that Descartes and other philosophers have argued about for centuries. Chemists, biologists, and neuroscientists join them in trying to explicate the mechanisms by which neurons communicate with each other and eventually form our thoughts, memories, emotions, and desires. At the other end of the spectrum, economists study how we balance choices about limited natural and financial resources, and anthropologists study the influence of culture on thought and the formation of societies. So at one end we find scientists studying atoms and cells, at the other end there are scientists studying entire groups of people. Cognitive psychologists tend to study the individual, and mental *systems* within individual brains, although ideally we try to stay informed of what our colleagues are doing. So cognition is a truly interdisciplinary endeavor, and this collection of readings is intended to reflect that.

Why Not a Textbook?

This book grew out of a course I took at the Massachusetts Institute of Technology (MIT) in 1975, from Susan Carey and Merrill Garrett (with occasional guest lectures by Mary Potter), and courses I taught at the University of Ore-

gon, Stanford University, and the University of California at Berkeley. When I took cognition at MIT, there were only two textbooks about cognition as a field (if it could even be thought of as a field then): Ulric Neisser's *Cognitive Psychology* and Michael Posner's *Cognition: An Introduction*. Professors Carey and Garrett supplemented these texts with a thick book of hand-picked readings from *Scientific American* and mainstream psychology journals. Reading journal articles prepared the students for the *debates* that characterize science. Susan and Merrill skillfully brought these debates out in the classroom, through interactive lectures and the Socratic method. Cognition is full of opposing theories and controversies. It is an empirical science, but in many cases the same data are used to support different arguments, and the reader must draw his or her own conclusions. The field of cognition is alive, dynamic, and rediscovering itself all the time. We should expect nothing less of the science devoted to understanding the mind.

Today there are many excellent textbooks and readers devoted to cognition. Textbooks are valuable because they select and organize a daunting amount of information and cover the essential points of a topic. The disadvantage is that they do not reflect how psychologists learn about new research—this is most often done through journal articles or "high-level" book chapters directed to the working researcher. More technical in nature, these sources typically reveal details of an experiment's design, the measures used, and how the findings are interpreted. They also reveal some of the inherent *ambiguity* in research (often hidden in a textbook's tidy summary). Frequently students, when confronted with the actual data of a study, find alternate interpretations of the findings, and come to discover firsthand that researchers are often forced to draw their own conclusions. By the time undergraduates take a course in cognition (usually their second or third course in psychology) they find themselves wondering if they ought to *major* in psychology, and a few even think about going to graduate school. I believe they ought to know more about what it is like to read actual psychology articles, so they'll know what they're getting into.

On the other hand, a book of readings composed exclusively of such primary sources would be difficult to read without a suitable grounding in the field and would leave out many important concepts, lacking an overview. That is, it might tend to emphasize the trees at the expense of the forest.

Therefore, the goal of this anthology is to combine the best of both kinds of readings. By compiling an anthology such as this, I was able to pick and choose my favorite articles, by experts on each topic. Of the thirty-nine selections, ten are from undergraduate textbooks, six are from professional journals, sixteen are chapters from "high-level" books aimed at advanced students and research scientists, and seven are more or less hybrids, coming from sources written for the educated layperson, such as *Scientific American* or popular books (e.g., Gardner, Norman). This book is *not* intended to be a collection of the most important papers in the history of cognitive psychology; other authors have done this extremely well, especially Lloyd Komatsu in his excellent *Experimenting with the Mind* (1994, Brooks/Cole). It is intended as a collection of readings that can serve as the principal text for a course in cognitive psychology or cognitive science.

The particular readings included here owe their evolution to a course I taught at the University of California at Berkeley in the fall of 1999, "Fundamental Issues in Cognitive Science." The readings for that course had been carefully honed over ten years by Stephen Palmer and Alison Gopnik, outstanding teachers whose courses are motivated by an understanding of the philosophical basis for contemporary cognitive psychology. I had never seen cognitive psychology taught this way, but once I did I couldn't imagine teaching it any other way. A fundamental assumption I share with them is that cognitive psychology is in many respects *empirical philosophy*. By that I mean that the core questions in cognitive psychology were for centuries considered the domain of philosophers. Some of these questions include: What is the nature of thought? Does language influence thought? Are memories and perceptions accurate? How can we ever know if other people are conscious?

Aristotle was the first information-processing theorist, and without exaggeration one can argue that modern cognitive psychology owes him its heritage. Descartes launched modern approaches to these questions, and much current debate references his work. But for Aristotle, Descartes, Hume, Locke, Husserl, and others, the questions remained in the realm of philosophy. A century and a half ago this all changed when Wundt, Fechner, Helmholtz, and their cohorts established the first laboratories in which they employed *empirical* methods to probe what had previously been impenetrable to true science: the mind. Philosophers framed the questions, and mental scientists (as they were then sometimes called) conducted experiments to answer them.

Today, the empirical work that interests me most in the field of Cognition is theory-driven and builds on these philosophical foundations. And a new group of philosophers, philosophers of mind, closely monitor the progress made by cognitive psychologists in order to interpret and debate their findings and to place them in a larger context.

Who Is This For?

The book you have before you is intended to be used as a text for the undergraduate cognitive psychology class I teach at McGill University. I hope that others will find some value in it as well. It should also be suitable for students who wish to acquaint themselves through self-study with important ideas in cognition. The ambitious student or professor may want to use this to supplement a regular textbook as a way to add other perspectives on the topics covered. It may also be of use to researchers as a resource that gathers up key articles in one place. It presupposes a solid background in introductory psychology and research methods. Students should have encountered most of these topics previously, and this book gives them an opportunity to explore them more deeply.

How the Book Is Organized and How It Differs from Other Books

The articles in this reader are organized thematically around topics traditionally found in a course on cognitive psychology or cognitive science at the uni-

versity level. The order of the readings could certainly be varied without loss of coherence, although I think that the first few readings fit better at the beginning. After that any order should work.

The readings begin with philosophical foundations, and it is useful to keep these in mind when reading the remainder of the articles. This reflects the view that good science builds on earlier foundations, even if it ultimately rejects them.

This anthology differs from most other cognition readers in its coverage of several topics not typically taught in cognition courses. One is human factors and ergonomics, the study of how we interact with tools, machines, and artifacts, and what cognitive psychology can tell us about how to improve the design of such objects (including computers); this is represented in the excellent papers by Don Norman. Another traditionally underrepresented topic, evolutionary psychology, is represented here by two articles, one by David Buss and his colleagues, and the other by John Tooby and Leda Cosmides. Also unusual are the inclusion of sections on music cognition, experimental design, and as mentioned before, philosophical foundations. You will find that there is somewhat *less* coverage of neuroscience and computer science perspectives on cognition, simply because in our department at McGill, we teach separate courses on those topics, and this reader reflects an attempt to reduce overlap.

Acknowledgments

I would like to thank the many publishers and authors who agreed to let their works be included here, my students, and Amy Brand, Tom Stone, Carolyn Anderson, Margy Avery, and Kathleen Caruso at MIT Press. I am indebted in particular to the following students from my cognition class for their tireless efforts at proofreading and indexing this book: Lindsay Ball, Ioana Dalca, Nora Hussein, Christine Kwong, Aliza Miller, Bianca Mugyenyi, Patrick Sabourin, and Hannah Weinstangel. I also would like to thank my wife, Caroline Traube, who is a constant source of surprise and inspiration and whose intuitions about cognitive psychology have led to many new studies. Finally, I was extraordinarily lucky to have three outstanding scholars as teachers: Mike Posner, Doug Hintzman, and Roger Shepard, to whom this book is dedicated. I would like to thank them for their patience, inspiration, support, and friendship.

PART I

Foundations—Philosophical Basis, The Mind/Body Problem

Chapter 1

Visual Awareness

Stephen E. Palmer

1.1 Philosophical Foundations

The first work on virtually all scientific problems was done by philosophers, and the nature of human consciousness is no exception. The issues they raised have framed the discussion for modern theories of awareness. Philosophical treatments of consciousness have primarily concerned two issues that we will discuss before considering empirical facts and theoretical proposals: The *mind-body problem* concerns the relation between mental events and physical events in the brain, and the *problem of other minds* concerns how people come to believe that other people (or animals) are also conscious.

1.1.1 The Mind-Body Problem

Although there is a long history to how philosophers have viewed the nature of the mind (sometimes equated with the soul), the single most important issue concerns what has come to be called the *mind-body problem*: What is the relation between mental events (e.g., perceptions, pains, hopes, desires, beliefs) and physical events (e.g., brain activity)? The idea that there is a mind-body problem to begin with presupposes one of the most important philosophical positions about the nature of mind. It is known as *dualism* because it proposes that mind and body are two different kinds of entities. After all, if there were no fundamental differences between mental and physical events, there would be no problem in saying how they relate to each other.

Dualism The historical roots of dualism are closely associated with the writings of the great French philosopher, mathematician, and scientist René Descartes. Indeed, the classical version of dualism, *substance dualism*, in which mind and body are conceived as two different substances, is often called *Cartesian dualism*. Because most philosophers find the notion of physical substances unproblematic, the central issue in philosophical debates over substance dualism is whether mental substances exist and, if so, what their nature might be. Vivid sensory experiences, such as the appearance of redness or the feeling of pain, are among the clearest examples, but substance dualists also include more abstract mental states and events such as hopes, desires, and beliefs.

The hypothesized mental substances are proposed to differ from physical ones in their fundamental properties. For example, all ordinary physical matter

From chapter 13 in *Vision Science: Photons to Phenomenology* (Cambridge, MA: MIT Press, 1999), 618–630. Reprinted with permission.

has a well-defined position, occupies a particular volume, has a definite shape, and has a specific mass. Conscious experiences, such as perceptions, remembrances, beliefs, hopes, and desires, do not appear to have readily identifiable positions, volumes, shapes, and masses. In the case of vision, however, one might object that visual experiences *do* have physical locations and extensions. There is an important sense in which my perception of a red ball on the table is located on the table where the ball is and is extended over the spherical volume occupied by the ball. What could be more obvious? But a substance dualist would counter that these are properties of the physical object that I perceive rather than properties of my perceptual experience itself. The experience is in my mind rather than out there in the physical environment, and the location, extension, and mass of these mental entities are difficult to define—unless one makes the problematic move of simply identifying them with the location, extension, and mass of my brain. Substance dualists reject this possibility, believing instead that mental states, such as perceptions, beliefs, and desires, are simply undefined with respect to position, extension, and mass. In this case, it makes sense to distinguish mental substances from physical ones on the grounds that they have fundamentally different properties.

We can also look at the issue of fundamental properties the other way around: Do experiences have any properties that ordinary physical matter does not? Two possibilities merit consideration. One is that experiences are *subjective phenomena* in the sense that they cannot be observed by anyone but the person having them. Ordinary matter and events, in contrast, are *objective phenomena* because they can be observed by anyone, at least in principle. The other is that experiences have what philosophers call *intentionality*: They inherently refer to things other than themselves.[1] Your experience of a book in front of you right now is about the book in the external world even though it arises from activity in your brain. This *directedness* of visual experiences is the source of the confusion we mentioned in the previous paragraph about whether your perceptions have location, extension, and so forth. The physical objects to which such perceptual experiences refer have these physical properties, but the experiences themselves do not. Intentionality does not seem to be a property that is shared by ordinary matter, and if this is true, it provides further evidence that conscious experience is fundamentally different.

It is possible to maintain a dualistic position and yet deny the existence of any separate mental substances, however. One can instead postulate that the brain has certain unique properties that constitute its mental phenomena. These properties are just the sorts of experiences we have as we go about our everyday lives, including perceptions, pains, desires, and thoughts. This philosophical position on the mind-body problems is called *property dualism*. It is a form of dualism because these properties are taken to be nonphysical in the sense of not being reducible to any standard physical properties. It is as though the physical brain contains some strange nonphysical features or dimensions that are qualitatively distinct from all physical features or dimensions.

These mental features or dimensions are usually claimed to be *emergent properties*: attributes that simply do not arise in ordinary matter unless it reaches a certain level or type of complexity. This complexity is certainly achieved in the human brain and may also be achieved in the brains of certain other animals.

The situation is perhaps best understood by analogy to the emergent property of being alive. Ordinary matter manifests this property only when it is organized in such a way that it is able to replicate itself and carry on the required biological processes. The difference, of course, is that being alive is a property that we can now explain in terms of purely physical processes. Property dualists believe that this will never be the case for mental properties.

Even if one accepts a dualistic position that the mental and physical are somehow qualitatively distinct, there are several different relations they might have to one another. These differences form the basis for several varieties of dualism. One critical issue is the direction of causation: Does it run from mind to brain, from brain to mind, or both? Descartes's position was that both sorts of causation are in effect: events in the brain can affect mental events, and mental events can also affect events in the brain. This position is often called *interactionism* because it claims that the mental and physical worlds can interact causally with each other in both directions. It seems sensible enough at an intuitive level. No self-respecting dualist doubts the overwhelming evidence that physical events in the brain cause the mental events of conscious experience. The pain that you feel in your toe, for example, is actually caused by the firing of neurons in your brain. Convincing evidence of this is provided by so-called *phantom limb pain*, in which amputees feel pain—sometimes excruciating pain—in their missing limbs (Chronholm, 1951; Ramachandran, 1996).

In the other direction, the evidence that mental events can cause physical ones is decidedly more impressionistic but intuitively satisfying to most interactionists. They point to the fact that certain mental events, such as my having the intention of raising my arm, appear to cause corresponding physical events, such as the raising of my arm—provided I am not paralyzed and my arm is not restrained in any way. The nature of this causation is scientifically problematic, however, because all currently known forms of causation concern physical events causing other physical events. Even so, other forms of causation that have not yet been identified may nevertheless exist.

Not all dualists are interactionists, however. An important alternative version of dualism, called *epiphenomenalism*, recognizes mental entities as being different in kind from physical ones yet denies that mental states play any causal role in the unfolding of physical events. An epiphenomenalist would argue that mental states, such as perceptions, intentions, beliefs, hopes, and desires, are merely ineffectual side effects of the underlying causal neural events that take place in our brains. To get a clearer idea of what this might mean, consider the following analogy: Imagine that neurons glow slightly as they fire in a brain and that this glowing is somehow akin to conscious experiences. The pattern of glowing in and around the brain (i.e., the conscious experience) is clearly caused by the firing of neurons in the brain. Nobody would question that. But the neural glow would be causally ineffectual in the sense that it would not cause neurons to fire any differently than they would if they did not glow. Therefore, causation runs in only one direction, from physical to mental, in an epiphenomenalist account of the mind-body problem. Although this position denies any causal efficacy to mental events, it is still a form of dualism because it accepts the existence of the "glow" of consciousness and maintains that it is qualitatively distinct from the neural firings themselves.

Idealism Not all philosophical positions on the mind-body problem are dualistic. The opposing view is *monism*: the idea that there is really just one sort of stuff after all. Not surprisingly, there are two sorts of monist positions—*idealism* and *materialism*—one for each kind of stuff there might be. A monist who believes there to be no physical world, but only mental events, is called an idealist (from the "ideas" that populate the mental world). This has not been a very popular position in the history of philosophy, having been championed mainly by the British philosopher Bishop Berkeley.

The most significant problem for idealism is how to explain the commonality of different people's perceptions of the same physical events. If a fire engine races down the street with siren blaring and red lights flashing, everyone looks toward it, and they all see and hear pretty much the same physical events, albeit from different vantage points. How is this possible if there is no physical world that is responsible for their simultaneous perceptions of the sound and sight of the fire engine? One would have to propose some way in which the minds of the various witnesses happen to be hallucinating exactly corresponding events at exactly corresponding times. Berkeley's answer was that God was responsible for this grand coordination, but such claims have held little sway in modern scientific circles. Without a cogent scientific explanation of the commonality of shared experiences of the physical world, idealism has largely become an historical curiosity with no significant modern following.

Materialism The vast majority of monists believe that only physical entities exist. They are called materialists. In contrast to idealism, materialism is a very common view among modern philosophers and scientists. There are actually two distinct forms of materialism, which depend on what their adherents believe the ultimate status of mental entities will be once their true physical nature is discovered. One form, called *reductive materialism*, posits that mental events will ultimately be reduced to material events in much the same way that other successful reductions have occurred in science (e.g., Armstrong, 1968). This view is also called *mind-brain identity theory* because it assumes that mental events are actually equivalent to brain events and can be talked about more or less interchangeably, albeit with different levels of precision.

A good scientific example of what reductive materialists believe will occur when the mental is reduced to the physical is the reduction in physics of thermodynamic concepts concerning heat to statistical mechanics. The temperature of a gas in classical thermodynamics has been shown to be equivalent to the average kinetic energy of its molecules in statistical mechanics, thus replacing the qualitatively distinct thermodynamic concept of heat with the more general and basic concept of molecular motion. The concept of heat did not then disappear from scientific vocabulary: it remains a valid concept within many contexts. Rather, it was merely given a more accurate definition in terms of molecular motion at a more microscopic level of analysis. According to reductive materialists, then, mental concepts will ultimately be redefined in terms of brain states and events, but their equivalence will allow mental concepts to remain valid and scientifically useful even after their brain correlates are discovered. For example, it will still be valid to say, "John is hungry," rather than, "Such-and-such pattern of neural firing is occurring in John's lateral hypothalamus."

The other materialist position, called *eliminative materialism*, posits that at least some of our current concepts concerning mental states and events will eventually be eliminated from scientific vocabulary because they will be found to be simply invalid (e.g., Churchland, 1990). The scenario eliminative materialists envision is thus more radical than the simple translation scheme we just described for reductive materialism. Eliminative materialists believe that some of our present concepts about mental entities (perhaps including perceptual experiences as well as beliefs, hopes, desires, and so forth) are so fundamentally flawed that they will someday be entirely replaced by a scientifically accurate account that is expressed in terms of the underlying neural events. An appropriate analogy here would be the elimination of the now-discredited ideas of "vitalism" in biology: the view that what distinguishes living from nonliving things is the presence of a mysterious and qualitatively distinct force or substance that is present in living objects and absent in nonliving ones. The discovery of the biochemical reactions that cause the replication of DNA by completely normal physical means ultimately undercut any need for such mystical concepts, and so they were banished from scientific discussion, never to be seen again.

In the same spirit, eliminative materialists believe that some mental concepts, such as perceiving, thinking, desiring, and believing, will eventually be supplanted by discussion of the precise neurological events that underlie them. Scientists would then speak exclusively of the characteristic pattern of neural firings in the appropriate nuclei of the lateral hypothalamus and leave all talk about "being hungry" or "the desire to eat" to historians of science who study archaic and discredited curiosities of yesteryear. Even the general public would eventually come to think and talk in terms of these neuroscientific explanations for experiences, much as modern popular culture has begun to assimilate certain notions about DNA replication, gene splicing, cloning, and related concepts into movies, advertising, and language.

Behaviorism Another position on the mind-body problem is *philosophical behaviorism*: the view that the proper way to talk about mental events is in terms of the overt, observable movements (behaviors) in which an organism engages. Because objective behaviors are measurable, quantifiable aspects of the physical world, behaviorism is, strictly speaking, a kind of materialism. It provides such a different perspective, however, that it is best thought of as a distinct view. Behaviorists differ markedly from standard materialists in that they seek to reduce mental events to behavioral events or dispositions rather than to neurophysiological events. They shun neural explanations not because they disbelieve in the causal efficacy of neural events, but because they believe that behavior offers a higher and more appropriate level of analysis. The radical behaviorist movement pressed for nothing less than redefining the scientific study of mind as the scientific study of behavior. And for many years, they succeeded in changing the agenda of psychology.

The behaviorist movement began with the writings of psychologist John Watson (1913), who advocated a thoroughgoing purge of everything mental from psychology. He reasoned that what made intellectual inquiries scientific rather than humanistic or literary was that the empirical data and theoretical constructs on which they rest are objective. In the case of empirical observations,

objectivity means that, given a description of what was done in a particular experiment, any scientist could repeat it and obtain essentially the same results, at least within the limits of measurement error. By this criterion, introspective studies of the qualities of perceptual experience were unscientific because they were not objective. Two different people could perform the same experiment (using themselves as subjects, of course) and report different experiences. When this happened—and it did—there was no way to resolve disputes about who was right. Both could defend their own positions simply by appealing to their private and privileged knowledge of their own inner states. This move protected their claims but blocked meaningful scientific debate.

According to behaviorists, scientists should study the behavior of organisms in a well-defined task situation. For example, rather than introspect about the nature of the perception of length, behaviorists would perform an experiment. Observers could be asked to discriminate which of two lines was longer, and their performance could be measured in terms of percentages of correct and incorrect responses for each pair of lines. Such an objective, behaviorally defined experiment could easily be repeated in any laboratory with different subjects to verify the accuracy and generality of its results. Watson's promotion of objective, behaviorally defined experimental methods—called *methodological behaviorism*—was a great success and strongly shaped the future of psychological research.

Of more relevance to the philosophical issue of the relation between mind and body, however, were the implications of the behaviorist push for objectivity in theoretical constructs concerning the mind. It effectively ruled out references to mental states and processes, replacing them with statements about an organism's propensity to engage in certain behaviors under certain conditions. This position is often called theoretical behaviorism or philosophical behaviorism. Instead of saying, "John is hungry," for example, which openly refers to a conscious mental experience (hunger) with which everyone is presumably familiar, a theoretical behaviorist would say something like "John has a propensity to engage in eating behavior in the presence of food." This propensity can be measured in a variety of objective ways—such as the amount of a certain food eaten when it was available after a certain number of hours since the last previous meal—precisely because it is about observable behavior.

But the behaviorist attempt to avoid talking about conscious experience runs into trouble when one considers all the conditions in which John might fail to engage in eating behavior even though he was hungry and food was readily available. Perhaps he could not see the food, for example, or maybe he was fasting. He might even have believed that the food was poisoned. It might seem that such conditions could be blocked simply by inserting appropriate provisions into the behavioral statement, such as "John had a propensity to engage in eating behavior in the presence of food, provided he perceived it, was not fasting, and did not believe it was poisoned." This move ultimately fails, however, for at least two reasons:

1. *Inability to enumerate all conditionals.* Once one begins to think of conditions that would have to be added to statements about behavioral dispositions, it quickly becomes apparent that there are indefinitely many.

Perhaps John fails to eat because his hands are temporarily paralyzed, because he has been influenced by a hypnotic suggestion, or whatever. This problem undercuts the claim that behavioral analyses of mental states are elegant and insightful, suggesting instead that they are fatally flawed or at least on the wrong track.

2. *Inability to eliminate mental entities*. The other problem is that the conditionals that must be enumerated frequently make reference to just the sorts of mental events that are supposed to be avoided. For example, whether John *sees* the food or not, whether he *intends* to fast, and what he *believes* about its being poisoned are all mentalistic concepts that have now been introduced into the supposedly behavioral definition. The amended version is therefore unacceptable to a strict theoretical behaviorist.

For such reasons, theoretical behaviorism ultimately failed. The problem, in a nutshell, was that behaviorists mistook the *epistemic status* of mental states (how we come to know about mental states in other people) for the *ontological status* of mental states (what their inherent nature is) (Searle, 1992). That is, we surely come to know about other people's mental states through their behavior, but this does not mean that the nature of these mental states is inherently behavioral.

Functionalism *Functionalism* was a movement in the philosophy of mind that began in the 1960s in close association with the earliest stirrings of cognitive science (e.g., Putnam, 1960). Its main idea is that a given mental state can be defined in terms of the causal relations that exist among that mental state, environmental conditions (inputs), organismic behaviors (outputs), and other mental states. Note that this is very much like behaviorism, but with the important addition of allowing other mental states into the picture. This addition enables a functionalist definition of hunger, for example, to refer to a variety of other mental states, such as perceptions, intentions, and beliefs, as suggested above. Functionalists are not trying to explain away mental phenomena as actually being propensities to behave in certain ways, as behaviorists did. Rather, they are trying to define mental states in terms of their relations to other mental states as well as to input stimuli and output behaviors. The picture that emerges is very much like information processing analyses. This is not surprising because functionalism is the philosophical foundation of modern computational theories of mind.

Functionalists aspired to more than just the overthrow of theoretical behaviorism, however. They also attempted to block reductive materialism by suggesting new criticisms of mind-brain identity theory. The basis of this criticism lies in the notion of *multiple realizability*: the fact that many different physical devices can serve the same function, provided they causally connect inputs and outputs in the same way via internal states (Putnam, 1967). For example, there are many different ways of building a thermostat. They all have the same function—to control the temperature in the thermostat's environment—but they realize it through very different physical implementations.

Multiple realizability poses the following challenge to identity theory. Suppose there were creatures from some other galaxy whose biology was based on silicon molecules rather than on carbon molecules, as ours is. Let us also

suppose that they were alive (even though the basis of their life was not DNA, but some functionally similar self-replicating molecule) and that they even look like people. And suppose further not only that their brains were constructed of elements that are functionally similar to neurons, but also that these elements were interconnected in just the way that neurons in our brains are. Indeed, their brains would be functionally isomorphic to ours, even though they were made of physically different stuff.

Functionalists then claim that these alien creatures would have the same mental states as we do—that is, the same perceptions, pains, desires, beliefs, and so on that populate our own conscious mental lives—provided that their internal states were analogously related to each other, to the external world, and to their behavior. This same approach can be generalized to argue for the possibility that computers and robots of the appropriate sort would also be conscious. Suppose, for example, that each neuron in a brain was replaced with a microcomputer chip that exactly simulated its firing patterns in response to all the neuron chips that provide its input. The computer that was thus constructed would fulfill the functionalist requirements for having the same mental states as the person whose brain was "electronically cloned." You should decide for yourself whether you believe that such a computer would actually have mental states or would merely act as though it had mental states. Once you have done so, try to figure out what criteria you used to decide. (For two contradictory philosophical views of this thought experiment, the reader is referred to Dennett (1991) and Searle (1993).)

Multiple realizability is closely related to differences between the algorithmic and implementation levels. The algorithmic level corresponds roughly to the functional description of the organism in terms of the relations among its internal states, its input information, and its output behavior. The implementation level corresponds to its actual physical construction. The functionalist notion of multiple realizability thus implies that there could be many different kinds of creatures that would have the same mental states as people do, at least defined in this way. If true, this would undercut identity theory, since mental events could not then be simply equated with particular neurological events; they would have to be equated with some more general class of physical events that would include, among others, silicon-based aliens and electronic brains.

The argument from multiple realizability is crucial to the functionalist theory of mind. Before we get carried away with the implications of multiple realizability, though, we must ask ourselves whether it is true or even remotely likely to be true. There is not much point in basing our understanding of consciousness on a functionalist foundation unless that foundation is well grounded. Is it? More important, how would we know if it were? We will address this topic shortly when we consider the problem of other minds.

Supervenience There is certainly some logical relation between brain activity and mental states such as consciousness, but precisely what it is has obviously been difficult to determine. Philosophers of mind have spent hundreds of years trying to figure out what it is and have spilled oceans of ink attacking and defending different positions. Recently, however, philosopher Jaegwon Kim (1978, 1993) has formulated a position with which most philosophers of mind

have been able to agree. This relation, called *supervenience*, is that any difference in conscious events requires some corresponding difference in underlying neural activity. In other words, mental events supervene on neural events because no two possible situations can be identical with respect to their neural properties while differing in their mental properties. It is a surprisingly weak relation, but it is better than nothing.

Supervenience does not imply that all differences in underlying neural activity result in differences in consciousness. Many neural events are entirely outside awareness, including those that control basic bodily functions such as maintaining gravitational balance and regulating heartbeat. But supervenience claims that no changes in consciousness can take place without some change in neural activity. The real trick, of course, is saying precisely what kinds of changes in neural events produce what kinds of changes in awareness.

1.1.2 The Problem of Other Minds

The functionalist arguments about multiple realizability are merely thought experiments because neither aliens nor electronic brains are currently at hand. Even so, the question of whether or not someone or something is conscious is central to the enterprise of cognitive science because the validity of such arguments rests on the answer. Formulating adequate criteria for consciousness is one of the thorniest problems in all of science. How could one possibly decide?

Asking how to discriminate conscious from nonconscious beings brings us face to face with another classic topic in the philosophy of mind: the *problem of other minds*. The issue at stake is how I know whether another creature (or machine) has conscious experiences. Notice that I did not say "how *we* know whether another creature has conscious experiences," because, strictly speaking, I do not know whether *you* do or not. This is because one of the most peculiar and unique features of my consciousness is its internal, private nature: Only I have direct access to my conscious experiences, and I have direct access only to my own. As a result, my beliefs that other people also have conscious experiences—and your belief that I do—appear to be inferences. Similarly, I may believe that dogs and cats, or even frogs and worms, are conscious. But in every case, the epistemological basis of my belief about the consciousness of other creatures is fundamentally different from knowledge of my own consciousness: I have direct access to my own experience and nobody else's.

Criteria for Consciousness If our beliefs that other people—and perhaps many animals as well—have experiences like ours are inferences, on what might such inferences be based? There seem to be at least two criteria.

1. *Behavioral similarity*. Other people act in ways that are roughly similar to my own actions when I am having conscious experiences. When I experience pain on stubbing my toe, for example, I may wince, say "Ouch!" and hold my toe while hopping on my other foot. When other people do similar things under similar circumstances, I presume they are experiencing a feeling closely akin to my own pain. Dogs also behave in seemingly analogous ways in what appear to be analogous situations in which they might experience pain, and so I also attribute this mental state of being in pain to them. The case is less compelling for creatures like frogs and

worms because their behavior is less obviously analogous to our own, but many people firmly believe that their behavior indicates that they also have conscious experiences such as pain.

2. *Physical similarity*. Other people—and, to a lesser degree, various other species of animals—are similar to me in their basic biological and physical structure. Although no two people are exactly the same, humans are generally quite similar to each other in terms of their essential biological constituents. We are all made of the same kind of flesh, blood, bone, and so forth, and we have roughly the same kinds of sensory organs. Many other animals also appear to be made of similar stuff, although they are morphologically different to varying degrees. Such similarities and differences may enter into our judgments of the likelihood that other creatures also have conscious experiences.

Neither condition alone is sufficient for a convincing belief in the reality of mental states in another creature. Behavioral similarity alone is insufficient because of the logical possibility of *automatons*: robots that are able to simulate every aspect of human behavior but have no experiences whatsoever. We may think that such a machine acts as if it had conscious experiences, but it could conceivably do so without actually having them. (Some theorists reject this possibility, however [e.g., Dennett, 1991].) Physical similarity alone is insufficient because we do not believe that even another living person is having conscious experiences when they are comatose or in a dreamless sleep. Only the two together are convincing. Even when both are present to a high degree, I still have no guarantee that such an inference is warranted. I only know that I myself have conscious experiences.

But what then is the status of the functionalist argument that an alien creature based on silicon rather than carbon molecules would have mental states like ours? This thought experiment is perhaps more convincing than the electronic-brained automaton because we have presumed that the alien is at least alive, albeit using some other physical mechanism to achieve this state of being. But logically, it would surely be unprovable that such silicon people would have mental states like ours, even if they acted very much the same and appeared very similar to people. In fact, the argument for functionalism from multiple realizability is no stronger than our intuitions that such creatures would be conscious. The strength of such intuitions can (and does) vary widely from one person to another.

The Inverted Spectrum Argument We have gotten rather far afield from visual perception in all this talk of robots, aliens, dogs, and worms having pains, but the same kinds of issues arise for perception. One of the classic arguments related to the problem of other minds—called the *inverted spectrum argument*—concerns the perceptual experience of color (Locke, 1690/1987). It goes like this: Suppose you grant that I have visual awareness in some form that includes differentiated experiences in response to different physical spectra of light (i.e., differentiated color perceptions). How can we know whether my color experiences are the same as yours?

The inverted spectrum argument refers to the possibility that my color experiences are exactly like your own, except for being spectrally inverted. In its

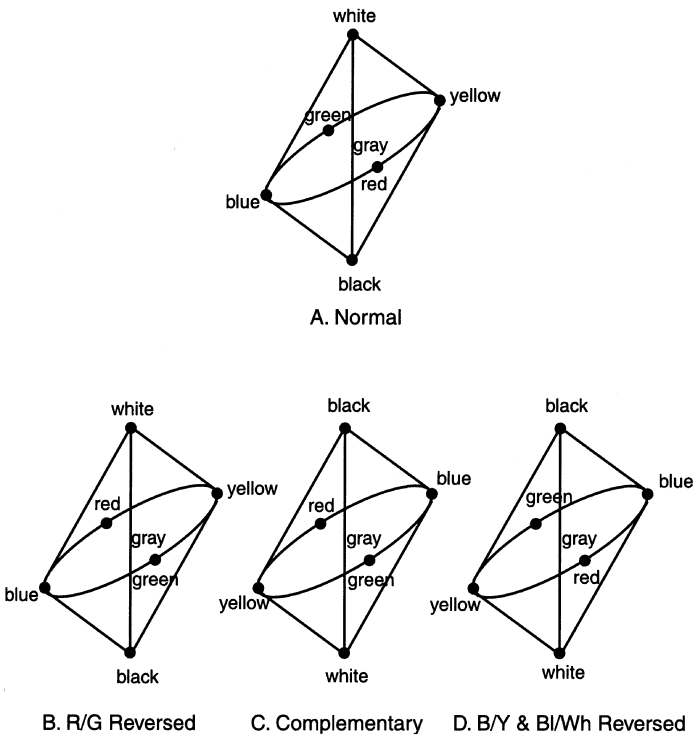


Figure 1.1
Sophisticated versions of the inverted spectrum argument. Transformations of the normal color solid (A) that would not be detectable by behavioral methods include (B) red-green reversal, which reflects each color about the blue-yellow-black-white place; (C) the complementary transformation, which reflects each color through the central point; and (D) blue-yellow and black-white reversal, which is the combination of both the two other transformations (B and C). (After Palmer, 1999.)

literal form, the inversion refers to reversing the mapping between color experiences and the physical spectrum of wavelengths of light, as though the rainbow had simply been reversed, red for violet (and vice versa) with everything in between being reversed in like manner. The claim of the inverted spectrum argument is that no one would ever be able to tell that you and I have different color experiences.

This particular form of color transformation would not actually work as intended because of the shape of the color solid (Palmer, 1999). The color solid is asymmetrical in that the most saturated blues and violets are darker than the most saturated reds and greens, which, in turn, are darker than the most saturated yellows and oranges (see figure 1.1A). The problem this causes for the literal inverted spectrum argument is that if my hues were simply reversed, your experience of yellow would be the same as my experience of blue-green, and so you would judge yellow to be darker than blue-green, whereas I would do the reverse. This difference would allow the spectral inversion of my color experiences (relative to yours) to be detected.

This problem may be overcome by using more sophisticated versions of the same color transformation argument (Palmer, 1999). The most plausible is

red-green reversal, in which my color space is the same as yours except for reflection about the blue-yellow plane, thus reversing reds and greens (see figure 1.1B). It does not suffer from problems concerning the differential lightness of blues and yellows because my blues correspond to your blues and my yellows to your yellows. Our particular shades of blues and yellows would be different—my greenish yellows and greenish blues would correspond to your reddish yellows (oranges) and reddish blues (purples), respectively, and vice versa—but gross differences in lightness would not be a problem.

There are other candidates for behaviorally undetectable color transformations as well (see figures 1.1C and 1.1D). The crucial idea in all these versions of the inverted spectrum argument is that if the color solid were symmetric with respect to some transformation—and this is at least roughly true for the three cases illustrated in figures 1.1B–1.1D—there would be no way to tell the difference between my color experiences and yours simply from our behavior. In each case, I would name colors in just the same way as you would, because these names are only *mediated* by our own private experiences of color. It is the sameness of the physical spectra that ultimately causes them to be named consistently across people, not the sameness of the private experiences. I would also describe relations between colors in the same way as you would: that focal blue is darker than focal yellow, that lime green is yellower than emerald green, and so forth. In fact, if I were in a psychological experiment in which my task was to rate pairs of color for similarity or dissimilarity, I would make the same ratings you would. I would even pick out the same unique hues as you would—the "pure" shades of red, green, blue, and yellow—even though my internal experiences of them would be different from yours. It would be extremely difficult, if not impossible, to tell from my behavior with respect to color that I experience it differently than you do.[2]

I suggested that red-green reversal is the most plausible form of color transformation because a good biological argument can be made that there should be some very small number of seemingly normal trichromats who should be red-green reversed. The argument for such *pseudo-normal color perception* goes as follows (Nida-Rümelin, 1996). Normal trichromats have three different pigments in their three cone types (figure 1.2A). Some people are red-green color blind because they have a gene that causes their long-wavelength (L) cones to have the same pigment as their medium-wavelength (M) cones (figure 1.2B). Other people have a different form of red-green color blindness because they have a different gene that causes their M cones to have the same pigment as their L cones (figure 1.2C). In both cases, people with these genetic defects lose the ability to experience both red and green because the visual system codes both colors by taking the difference between the outputs of these two cone types. But suppose that someone had the genes for *both* of these forms of red-green color blindness. Their L cones would have the M pigment, and their M cones would have the L pigment (figure 1.2D). Such doubly color blind individuals would therefore not be red-green color blind at all, but red-green-reversed trichromats.[3] Statistically, they should be very rare (about 14 per 10,000 males), but they should exist. If they do, they are living proof that this color transformation is either undetectable or very difficult to detect by purely behavioral means, because nobody has ever detected one!

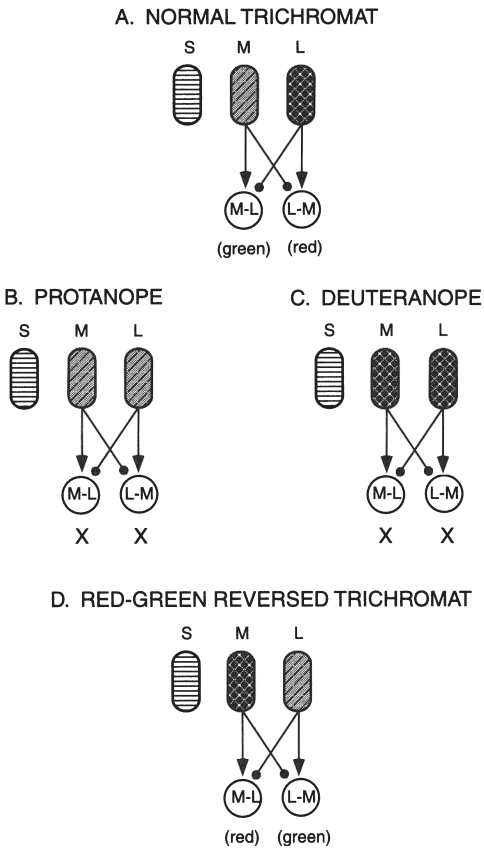


Figure 1.2
A biological basis for red-green-reversed trichromats. Normal trichromats have three different pigments in the retinal cones (A), whereas red-green color blind individuals have the same pigment in their L and M cones (B and C). People with the genes for both forms of red-green color blindness, however, would be red-green-reversed trichromats (D).

These color transformation arguments are telling criticisms against the completeness of any definition of conscious experience based purely on behavior. Their force lies in the fact that there could be identical behavior in response to identical environmental stimulation without there being corresponding identical experiences underlying them, even if we grant that the other person has experiences to begin with.

Phenomenological Criteria Let us return to the issue of criteria for consciousness: How are we to tell whether a given creature is conscious or not? Clearly, phenomenological experience is key. In fact, it is the defining characteristic, the necessary and sufficient condition, for attributing consciousness to something. I know that I am conscious precisely because I have such experiences. This is often called *first-person knowledge* or *subjective knowledge* because it is available only to the self (i.e., the first-person or subject). In his classic essay "What Is It Like to Be a Bat?" philosopher Thomas Nagel (1974) identifies the

phenomenological position with what it is like to *be* some person, creature, or machine in a given situation. In the case of color perception, for example, it is what it is like for you to experience a particular shade of redness or pale blueness or whatever. This much seems perfectly clear. But if it is so clear, then why not simply define consciousness with respect to such phenomenological criteria?

As we said before, the difficulty is that first-person knowledge is available only to the self. This raises a problem for scientific explanations of consciousness because the scientific method requires its facts to be objective in the sense of being available to any scientist who undertakes the same experiment. In all matters except consciousness, this appears to work very well. But consciousness has the extremely peculiar and elusive property of being directly accessible only to the self, thus blocking the usual methods of scientific observation. Rather than observing consciousness itself in others, the scientist is forced to observe the correlates of consciousness, the "shadows of consciousness," as it were. Two sorts of shadows are possible to study: behavior and physiology. Neither is consciousness itself, but both are (or seem likely to be) closely related.

Behavioral Criteria The most obvious way to get an objective, scientific handle on consciousness is to study behavior, as dictated by methodological behaviorism. Behavior is clearly objective and observable in the third-person sense. But how is it related to consciousness? The link is the assumption that if someone or something behaves enough like I do, it must be conscious like I am. After all, I believe I behave in the ways I do because of my own conscious experiences, and so (presumably) do others. I wince when I am in pain, eat when I am hungry, and duck when I perceive a baseball hurtling toward my head. If I were comatose, I would not behave in any of these ways, even in the same physical situations.

Behavioral criteria for consciousness are closely associated with what is called *Turing's test*. This test was initially proposed by the brilliant mathematician Alan Turing (1950), inventor of the digital computer, to solve the problem of how to determine whether a computing machine could be called "intelligent." Wishing to avoid purely philosophical debates, Turing imagined an objective behavioral procedure for deciding the issue by setting up an *imitation game*. A person is seated at a computer terminal that allows her to communicate either with a real person or with a computer that has been programmed to behave intelligently (i.e., like a person). This interrogator's job is to decide whether she is communicating with a person or the computer. The terminal is used simply to keep the interrogator from using physical appearance as a factor in the decision, since appearance presumably does not have any logical bearing on intelligence.

The interrogator is allowed to ask anything she wants. For example, she could ask the subject to play a game of chess, engage in a conversation on current events, or describe its favorite TV show. Nothing is out of bounds. She could even ask whether the subject is intelligent. A person would presumably reply affirmatively, but then so would a properly programmed computer. If the interrogator could not tell the difference between interacting with real people

and with the computer, Turing asserted that the computer should be judged "intelligent." It would then be said to have "passed Turing's test."

Note that Turing's test is a strictly behavioral test because the interrogator has no information about the physical attributes of the subject, but only about its behavior. In the original version, this behavior is strictly verbal, but there is no reason in principle why it needs to be restricted in this way. The interrogator could ask the subject to draw pictures or even to carry out tasks in the real world, provided the visual feedback the interrogator received did not provide information about the physical appearance of the subject.

The same imitation game can be used for deciding about the appropriateness of any other cognitive description, including whether the subject is "conscious." Again, simply asking the subject whether it is conscious will not discriminate between the machine and a person because the machine can easily be programmed to answer that question in the affirmative. Similarly, appropriate responses to questions asking it to describe the nature of its visual experiences or pain experiences could certainly be programmed. But even if they could, would that necessarily mean that the computer would *be* conscious or only that it would *act as if it were* conscious?

If one grants that physical appearance should be irrelevant to whether something is conscious or not, Turing's test seems to be a fair and objective procedure. But it also seems that there is a fact at issue here rather than just an opinion—namely, whether the target object is actually *conscious* or merely simulating consciousness—and Turing's test should stand or fall on whether it gives the correct answer. The problem is that it is not clear that it will. As critics readily point out, it cannot distinguish between a conscious entity and one that only acts as if it were conscious—an automaton or a zombie. To assert that Turing's test actually gives the correct answer to the factual question of consciousness, one must assume that it is impossible for something to act as if it is conscious without actually being so. This is a highly questionable assumption, although some have defended it (e.g., Dennett, 1991). If it is untrue, then passing Turing's test is not a sufficient condition for consciousness, because automatons can pass it without being conscious.

Turing's test also runs into trouble as a necessary condition for consciousness. The relevant question here is whether something can be conscious and still fail Turing's test. Although this might initially seem unlikely, consider a person who has an unusual medical condition that disables the use of all the muscles required for overt behavior yet keeps all other bodily functions intact, including all brain functions. This person would be unable to behave in any way yet would still be fully conscious when awake. Turing's test thus runs afoul as a criterion for consciousness because behavior's link to consciousness can be broken under unlikely but easily imaginable circumstances.

We appear to be on the horns of a dilemma with respect to the criteria for consciousness. Phenomenological criteria are valid by definition but do not appear to be scientific by the usual yardsticks. Behavioral criteria are scientific by definition but are not necessarily valid. The fact that scientists prefer to rely on respectable but possibly invalid behavioral methods brings to mind the streetlight parable: A woman comes upon a man searching for something under a streetlight at night. The man explains that he has lost his keys, and they both

search diligently for some time. The woman finally asks the man where he thinks he lost them, to which he replies, "Down the street in the middle of the block." When she then asks why he is looking here at the corner, he replies, "Because this is where the light is." The problem is that consciousness does not seem to be where behavioral science can shed much light on it.

Physiological Criteria Modern science has another card to play, however, and that is the biological substrate of consciousness. Even if behavioral methods cannot penetrate the subjectivity barrier of consciousness, perhaps physiological methods can. In truth, few important facts are yet known about the biological substrates of consciousness. There are not even very many hypotheses, although several speculations have recently been proposed (e.g., Baars, 1988; Crick, 1994; Crick & Koch, 1990, 1995, 1998; Edelman, 1989). Even so, it is possible to speculate about the promise such an enterprise might hold as a way of defining and theorizing about consciousness. It is important to remember that in doing so, we are whistling in the dark, however.

Let us suppose, just for the sake of argument, that neuroscientists discover some crucial feature of the neural activity that underlies consciousness. Perhaps all neural activity that gives rise to consciousness occurs in some particular layer of cerebral cortex, or in neural circuits that are mediated by some particular neurotransmitter, or in neurons that fire at a temporal spiking frequency of about 40 times per second. If something like one of these assertions were true—and, remember, we are just making up stories here—could we then define consciousness objectively in terms of that form of neural activity? If we could, would this definition then replace the subjective definition in terms of experience? And would such a biological definition then constitute a theory of consciousness?

The first important observation about such an enterprise is that biology cannot really give us an objective definition of consciousness independent of its subjective definition. The reason is that we need the subjective definition to determine what physiological events correspond to consciousness in the first place. Suppose we knew all of the relevant biological events that occur in human brains. We still could not provide a biological account of consciousness because we would have no way to tell which brain events were conscious and which ones were not. Without that crucial information, a biological definition of consciousness simply could not get off the ground. To determine the biological correlates of consciousness, one must be able to designate the events to which they are being correlated (i.e., conscious ones), and this requires a subjective definition.

For this reason, any biological definition of consciousness would always be derived from the subjective definition. To see this in a slightly different way, consider what would constitute evidence that a given biological definition was incorrect. If brain activity of type C were thought to define consciousness, it could be rejected for either of two reasons: if type C brain activity were found to result in nonconscious processing of some sort or if consciousness were found to occur in the absence of type C brain activity. The crucial observation for present purposes is that neither of these possibilities could be evaluated without an independent subjective definition of consciousness.

Correlational versus Causal Theories In considering the status of physiological statements about consciousness, it is important to distinguish two different sorts, which we will call *correlational* and *causal*. Correlational statements concern what type of physiological activity takes place when conscious experiences are occurring that fail to take place when they are not. Our hypothetical examples in terms of a specific cortical location, a particular neurotransmitter, or a particular rate of firing are good examples. The common feature of these hypotheses is that they are merely correlational: They only claim that the designated feature of brain activity is associated with consciousness; they don't explain why that association exists. In other words, they provide no causal analysis of how this particular kind of brain activity produces consciousness. For this reason they fail to fill the explanatory gap that we mentioned earlier. Correlational analyses merely designate a subset of neural activity in the brain according to some particular property with which consciousness is thought to be associated. No explanation is given for this association; it simply is the sort of activity that accompanies consciousness.

At this point we should contrast such correlational analyses with a good example of a causal one: an analysis that provides a scientifically plausible explanation of how a particular form of brain activity actually causes conscious experience. Unfortunately, no examples of such a theory are available. In fact, to this writer's knowledge, nobody has ever suggested a theory that the scientific community regards as giving even a remotely plausible causal account of how consciousness arises or why it has the particular qualities it does. This does not mean that such a theory is impossible in principle, but only that no serious candidate has been generated in the past several thousand years.

A related distinction between correlational and causal biological definitions of consciousness is that they would differ in generalizability. Correlational analyses would very likely be specific to the type of biological system within which they had been discovered. In the best-case scenario, a good correlational definition of human consciousness might generalize to chimpanzees, possibly even to dogs or rats, but probably not to frogs or snails because their brains are simply too different. If a correlational analysis showed that activity mediated by a particular neurotransmitter was the seat of human consciousness, for example, would that necessarily mean that creatures without that neurotransmitter were nonconscious? Or might some other evolutionarily related neural transmitter serve the same function in brains lacking that one? Even more drastically, what about extraterrestrial beings whose whole physical make-up might be radically different from our own? In such cases, a correlational analysis is almost bound to break down.

An adequate causal theory of consciousness might have a fighting chance, however, because the structure of the theory itself could provide the lines along which generalization would flow. Consider the analogy to a causal theory of life based on the structure of DNA. The analysis of how the double helical structure of DNA allows it to reproduce itself in an entirely mechanistic way suggests that biologists could determine whether alien beings were alive in the same sense as living organisms on earth by considering the nature of their molecular basis and its functional ability to replicate itself and to support the organism's lifelike functions. An alien object containing the very same set of

four component bases as DNA (adenine, guanine, thymine, and cytosine) in some very different global structure that did not allow self-replication would not be judged to be alive by such biological criteria, yet another object containing very different components in some analogous arrangement that allowed for self-replication might be. Needless to say, such an analysis is a long way off in the case of consciousness.

Notes

1. The reader is warned not to confuse intentionality with the concept of "intention" in ordinary language. Your intentions have intentionality in the sense that they may refer to things other than themselves—for example, your intention to feed your cat refers to your cat, its food, and yourself—but no more so than other mental states you might have, such as beliefs, desires, perceptions, and pains. The philosophical literature on the nature of intentionality is complex and extensive. The interested reader is referred to Bechtel (1988) for an overview of this topic.
2. One might think that if white and black were reversed, certain reflexive behaviors to light would somehow betray the difference. This is not necessarily the case, however. Whereas you would squint your eyes when you experienced intense brightness in response to bright sunlight, I would also squint my eyes in response to large amounts of sunlight. The only difference is that my experience of brightness under these conditions would be the same as your experience of darkness. It sounds strange, but I believe it would all work out properly.
3. One could object that the only thing that differentiates M and L cones is the pigment that they contain, so people with both forms of red-green color blindness would actually be normal trichromats rather than red-green-reversed ones. There are two other ways in which M and L cones might be differentiated, however. First, if the connections of M and L cones to other cells of the visual system are not completely symmetrical, they can be differentiated by these connections independently of their pigments. Second, they may be differentiable by their relation to the genetic codes that produced them.

References

Armstrong, D. M. (1968). *A materialist theory of the mind*. London: Routledge & Kegan Paul.

Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge, England: Cambridge University Press.

Churchland, P. M. (1990). Current eliminativism. In W. G. Lycan (Ed.), *Mind and cognition: A reader* (pp. 206–223). Oxford, England: Basil Blackwell.

Crick, F. H. C. (1994). *The astonishing hypothesis: The scientific search for the soul*. New York: Scribner.

Crick, F. H. C., & Koch, C. (1990). Toward a neurobiological theory of consciousness. *Seminars in the Neurosciences, 2*, 263–275.

Crick, F. H. C., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature, 375*, 121–123.

Crick, F. H. C., & Koch, C. (1998). Consciousness and neuroscience. *Cerebral cortex, 8*, 97–107.

Cronholm, B. (1951). Phantom limbs in amputees. *Acta Psychiatrica Scandinavica, 72* (Suppl.).

Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown.

Edelman, G. M. (1989). *The remembered present: A biological theory of consciousness*. New York: Basic Books.

Kim, J. (1978). Supervenience and nomological incommensurables. *American Philosophical Quarterly, 15*, 149–156.

Kim, J. (1993). *Supervenience and mind*. Cambridge, England: Cambridge University Press.

Locke, J. (1690/1987). *An essay concerning human understanding*. Oxford, England: Basil Blackwell.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*, 435–450.

Palmer, S. E. (1999). Color, consciousness, and the isomorphism constraint. *Behavioural and Brain Sciences, 22*(6), 923–989.

Putnam, H. (1960). Minds and machines. In S. Hook (Ed.), *Dimensions of mind*. New York: Collier Books.

Putnam, H. (1967). Psychological predicates. In W. Captain & D. Merrill (Eds.), *Art, mind, and religion* (pp. 35–48). Pittsburgh: University of Pittsburgh Press.

Ramachandran, V. S., Levi, L., Stone, L., Rogers-Ramachandran, D., McKinney, R., Stalcup, M., Arcilla, G., Sweifler, R., Schatz, A., Flippin, A. (1996). Illusions of body image: What they reveal about human nature. In R. R. Llinas and P. S. Churchland (Eds.), *The mind-brain continuum: Sensory processes* (pp. 29–60). Cambridge, MA: MIT Press.

Searle, J. R. (1992). *The rediscovery of mind*. Cambridge, MA: MIT Press.

Turing, S. (1959). *Alan M. Turing*. Cambridge, England: W. Heffer.

Chapter 2

Where Am I?

Daniel C. Dennett

Now that I've won my suit under the Freedom of Information Act, I am at liberty to reveal for the first time a curious episode in my life that may be of interest not only to those engaged in research in the philosophy of mind, artificial intelligence and neuroscience but also to the general public.

Several years ago I was approached by Pentagon officials who asked me to volunteer for a highly dangerous and secret mission. In collaboration with NASA and Howard Hughes, the Department of Defense was spending billions to develop a Supersonic Tunneling Underground Device, or STUD. It was supposed to tunnel through the earth's core at great speed and deliver a specially designed atomic warhead "right up the Red's missile silos," as one of the Pentagon brass put it.

The problem was that in an early test they had succeeded in lodging a warhead about a mile deep under Tulsa, Oklahoma, and they wanted me to retrieve it for them. "Why me?" I asked. Well, the mission involved some pioneering applications of current brain research, and they had heard of my interest in brains and of course my Faustian curiosity and great courage and so forth.... Well, how could I refuse? The difficulty that brought the Pentagon to my door was that the device I'd been asked to recover was fiercely radioactive, in a new way. According to monitoring instruments, something about the nature of the device and its complex interactions with pockets of material deep in the earth had produced radiation that could cause severe abnormalities in certain tissues of the brain. No way had been found to shield the brain from these deadly rays, which were apparently harmless to other tissues and organs of the body. So it had been decided that the person sent to recover the device should *leave his brain behind*. It would be kept in a safe place where it could execute its normal control functions by elaborate radio links. Would I submit to a surgical procedure that would completely remove my brain, which would then be placed in a life-support system at the Manned Spacecraft Center in Houston? Each input and output pathway, as it was severed, would be restored by a pair of microminiaturized radio transceivers, one attached precisely to the brain, the other to the nerve stumps in the empty cranium. No information would be lost, all the connectivity would be preserved. At first I was a bit reluctant. Would it really work? The Houston brain surgeons encouraged me. "Think of it," they said, "as a mere *stretching* of the nerves. If your brain were just moved over an

From chapter 17 in *Brainstorms* (Cambridge, MA: MIT Press, 1978), 310–323. Reprinted with permission.

inch in your skull, that would not alter or impair your mind. We're simply going to make the nerves indefinitely elastic by splicing radio links into them."

I was shown around the life-support lab in Houston and saw the sparkling new vat in which my brain would be placed, were I to agree. I met the large and brilliant support team of neurologists, hematologists, biophysicists, and electrical engineers, and after several days of discussions and demonstrations, I agreed to give it a try. I was subjected to an enormous array of blood tests, brain scans, experiments, interviews, and the like. They took down my autobiography at great length, recorded tedious lists of my beliefs, hopes, fears, and tastes. They even listed my favorite stereo recordings and gave me a crash session of psychoanalysis.

The day for surgery arrived at last and of course I was anesthetized and remember nothing of the operation itself. When I came out of anesthesia, I opened my eyes, looked around, and asked the inevitable, the traditional, the lamentably hackneyed post-operative question: "Where am I?" The nurse smiled down at me. "You're in Houston," she said, and I reflected that this still had a good chance of being the truth one way or another. She handed me a mirror. Sure enough, there were the tiny antennae poking up through their titanium ports cemented into my skull.

"I gather the operation was a success," I said, "I want to go see my brain." They led me (I was a bit dizzy and unsteady) down a long corridor and into the life-support lab. A cheer went up from the assembled support team, and I responded with what I hoped was a jaunty salute. Still feeling lightheaded, I was helped over to the life-support vat. I peered through the glass. There, floating in what looked like ginger-ale, was undeniably a human brain, though it was almost covered with printed circuit chips, plastic tubules, electrodes, and other paraphernalia. "Is that mine?" I asked. "Hit the output transmitter switch there on the side of the vat and see for yourself," the project director replied. I moved the switch to *off*, and immediately slumped, groggy and nauseated, into the arms of the technicians, one of whom kindly restored the switch to its *on* position. While I recovered my equilibrium and composure, I thought to myself: "Well, here I am, sitting on a folding chair, staring through a piece of plate glass at my own brain.... But wait," I said to myself, "shouldn't I have thought, 'Here I am, suspended in a bubbling fluid, being stared at by my own eyes'?" I tried to think this latter thought. I tried to project it into the tank, offering it hopefully to my brain, but I failed to carry off the exercise with any conviction. I tried again. "Here am *I*, Daniel Dennett, suspended in a bubbling fluid, being stared at by my own eyes." No, it just didn't work. Most puzzling and confusing. Being a philosopher of firm physicalist conviction, I believed unswervingly that the tokening of my thoughts was occurring somewhere in my brain: yet, when I thought "Here I am," where the thought occurred to me was *here*, outside the vat, where I, Dennett, was standing staring at my brain.

I tried and tried to think myself into the vat, but to no avail. I tried to build up to the task by doing mental exercises. I thought to myself, "The sun is shining *over there*," five times in rapid succession, each time mentally ostending a different place: in order, the sun-lit corner of the lab, the visible front lawn of the hospital, Houston, Mars, and Jupiter. I found I had little difficulty in getting my "there's" to hop all over the celestial map with their proper references. I

could loft a "there" in an instant through the farthest reaches of space, and then aim the next "there" with pinpoint accuracy at the upper left quadrant of a freckle on my arm. Why was I having such trouble with "here"? "Here in Houston" worked well enough, and so did "here in the lab," and even "here in this part of the lab," but "here in the vat" always seemed merely an unmeant mental mouthing. I tried closing my eyes while thinking it. This seemed to help, but still I couldn't manage to pull it off, except perhaps for a fleeting instant. I couldn't be sure. The discovery that I couldn't be sure was also unsettling. How did I know *where* I meant by "here" when I thought "here"? Could I *think* I meant one place when in fact I meant another? I didn't see how that could be admitted without untying the few bonds of intimacy between a person and his own mental life that had survived the onslaught of the brain scientists and philosophers, the physicalists and behaviorists. Perhaps I was incorrigible about where I *meant* when I said "here." But in my present circumstances it seemed that either I was doomed by sheer force of mental habit to thinking systematically false indexical thoughts, or where a person is (and hence where his thoughts are tokened for purposes of semantic analysis) is not necessarily where his brain, the physical seat of his soul, resides. Nagged by confusion, I attempted to orient myself by falling back on a favorite philosopher's ploy. I began naming things.

"Yorick," I said aloud to my brain, "you are my brain. The rest of my body, seated in this chair, I dub 'Hamlet.'" So here we all are: Yorick's my brain, Hamlet's my body, and I am Dennett. *Now*, where am I? And when I think "where am I?" where's that thought tokened? Is it tokened in my brain, lounging about in the vat, or right here between my ears where it *seems* to be tokened? Or nowhere? Its *temporal* coordinates give me no trouble; must it not have spatial coordinates as well? I began making a list of the alternatives.

1. *Where Hamlet goes, there goes Dennett*. This principle was easily refuted by appeal to the familiar brain transplant thought-experiments so enjoyed by philosophers. If Tom and Dick switch brains, Tom is the fellow with Dick's former body—just ask him; he'll claim to be Tom, and tell you the most intimate details of Tom's autobiography. It was clear enough, then, that my current body and I could part company, but not likely that I could be separated from my brain. The rule of thumb that emerged so plainly from the thought experiments was that in a brain-transplant operation, one wanted to be the *donor*, not the recipient. Better to call such an operation a *body*-transplant, in fact. So perhaps the truth was,

2. *Where Yorick goes, there goes Dennett*. This was not at all appealing, however. How could I be in the vat and not about to go anywhere, when I was so obviously outside the vat looking in and beginning to make guilty plans to return to my room for a substantial lunch? This begged the question I realized, but it still seemed to be getting at something important. Casting about for some support for my intuition, I hit upon a legalistic sort of argument that might have appealed to Locke.

Suppose, I argued to myself, I were now to fly to California, rob a bank, and be apprehended. In which state would I be tried: In California, where the robbery took place, or in Texas, where the brains of the outfit were located? Would I be a California felon with an out-of-state brain, or a Texas felon remotely

controlling an accomplice of sorts in California? It seemed possible that I might beat such a rap just on the undecidability of that jurisdictional question, though perhaps it would be deemed an inter-state, and hence Federal, offense. In any event, suppose I were convicted. Was it likely that California would be satisfied to throw Hamlet into the brig, knowing that Yorick was living the good life and luxuriously taking the waters in Texas? Would Texas incarcerate Yorick, leaving Hamlet free to take the next boat to Rio? This alternative appealed to me. Barring capital punishment or other cruel and unusual punishment, the state would be obliged to maintain the life-support system for Yorick though they might move him from Houston to Leavenworth, and aside from the unpleasantness of the opprobrium, I, for one, would not mind at all and would consider myself a free man under those circumstances. If the state has an interest in forcibly relocating persons in institutions, it would fail to relocate me in any institution by locating Yorick there. If this were true, it suggested a third alternative.

3. *Dennett is wherever he thinks he is*. Generalized, the claim was as follows: At any given time a person has a *point of view*, and the location of the point of view (which is determined internally by the content of the point of view) is also the location of the person.

Such a proposition is not without its perplexities, but to me it seemed a step in the right direction. The only trouble was that it seemed to place one in a heads-I-win/tails-you-lose situation of unlikely infallibility as regards location. Hadn't I myself often been wrong about where I was, and at least as often uncertain? Couldn't one get lost? Of course, but getting lost *geographically* is not the only way one might get lost. If one were lost in the woods one could attempt to reassure oneself with the consolation that at least one knew where one was: one was right *here* in the familiar surroundings of one's own body. Perhaps in this case one would not have drawn one's attention to much to be thankful for. Still, there were worse plights imaginable, and I wasn't sure I wasn't in such a plight right now.

Point of view clearly had something to do with personal location, but it was itself an unclear notion. It was obvious that the content of one's point of view was not the same as or determined by the content of one's beliefs or thoughts. For example, what should we say about the point of view of the Cinerama viewer who shrieks and twists in his seat as the roller-coaster footage overcomes his psychic distancing? Has he forgotten that he is safely seated in the theater? Here I was inclined to say that the person is experiencing an illusory shift in point of view. In other cases, my inclination to call such shifts illusory was less strong. The workers in laboratories and plants who handle dangerous materials by operating feedback-controlled mechanical arms and hands undergo a shift in point of view that is crisper and more pronounced than anything Cinerama can provoke. They can feel the heft and slipperiness of the containers they manipulate with their metal fingers. They know perfectly well where they are and are not fooled into false beliefs by the experience, yet it is as if they were inside the isolation chamber they are peering into. With mental effort, they can manage to shift their point of view back and forth, rather like making a transparent Neckar cube or an Escher drawing change orientation

before one's eyes. It does seem extravagant to suppose that in performing this bit of mental gymnastics, they are transporting *themselves* back and forth.

Still their example gave me hope. If I was in fact in the vat in spite of my intuitions, I might be able to train myself to adopt that point of view even as a matter of habit. I should dwell on images of myself comfortably floating in my vat, beaming volitions to that familiar body *out there*. I reflected that the ease or difficulty of this task was presumably independent of the truth about the location of one's brain. Had I been practicing before the operation, I might now be finding it second nature. You might now yourself try such a *tromp l'oeil*. Imagine you have written an inflammatory letter which has been published in the *Times*, the result of which is that the Government has chosen to impound your brain for a probationary period of three years in its Dangerous Brain Clinic in Bethesda, Maryland. Your body of course is allowed freedom to earn a salary and thus to continue its function of laying up income to be taxed. At this moment, however, your body is seated in an auditorium listening to a peculiar account by Daniel Dennett of his own similar experience. Try it. Think yourself to Bethesda, and then hark back longingly to your body, far away, and yet *seeming* so near. It is only with long-distance restraint (yours? the Government's?) that you can control your impulse to get those hands clapping in polite applause before navigating the old body to the rest room and a well-deserved glass of evening sherry in the lounge. The task of imagination is certainly difficult, but if you achieve your goal the results might be consoling.

Anyway, there I was in Houston, lost in thought as one might say, but not for long. My speculations were soon interrupted by the Houston doctors, who wished to test out my new prosthetic nervous system before sending me off on my hazardous mission. As I mentioned before, I was a bit dizzy at first, and not surprisingly, although I soon habituated myself to my new circumstances (which were, after all, well nigh indistinguishable from my old circumstances). My accommodation was not perfect, however, and to this day I continue to be plagued by minor coordination difficulties. The speed of light is fast, but finite, and as my brain and body move farther and farther apart, the delicate interaction of my feedback systems is thrown into disarray by the time lags. Just as one is rendered close to speechless by a delayed or echoic hearing of one's speaking voice so, for instance, I am virtually unable to track a moving object with my eyes whenever my brain and my body are more than a few miles apart. In most matters my impairment is scarcely detectable, though I can no longer hit a slow curve ball with the authority of yore. There are some compensations of course. Though liquor tastes as good as ever, and warms my gullet while corroding my liver, I can drink it in any quantity I please, without becoming the slightest bit inebriated, a curiosity some of my close friends may have noticed (though I occasionally have *feigned* inebriation, so as not to draw attention to my unusual circumstances). For similar reasons, I take aspirin orally for a sprained wrist, but if the pain persists I ask Houston to administer codeine to me *in vitro*. In times of illness the phone bill can be staggering.

But to return to my adventure. At length, both the doctors and I were satisfied that I was ready to undertake my subterranean mission. And so I left my brain in Houston and headed by helicopter for Tulsa. Well, in any case, that's

the way it seemed to me. That's how I would put it, just off the top of my head as it were. On the trip I reflected further about my earlier anxieties and decided that my first post-operative speculations had been tinged with panic. The matter was not nearly as strange or metaphysical as I had been supposing. Where was I? In two places, clearly: both inside the vat and outside it. Just as one can stand with one foot in Connecticut and the other in Rhode Island, I was in two places at once. I had become one of those scattered individuals we used to hear so much about. The more I considered this answer, the more obviously true it appeared. But, strange to say, the more true it appeared, the less important the question to which it could be the true answer seemed. A sad, but not unprecedented, fate for a philosophical question to suffer. This answer did not completely satisfy me, of course. There lingered some question to which I should have liked an answer, which was neither "Where are all my various and sundry parts?" nor "What is my current point of view?" Or at least there seemed to be such a question. For it did seem undeniable that in some sense *I* and not merely *most of me* was descending into the earth under Tulsa in search of an atomic warhead.

When I found the warhead, I was certainly glad I had left my brain behind, for the pointer on the specially built Geiger counter I had brought with me was off the dial. I called Houston on my ordinary radio and told the operation control center of my position and my progress. In return, they gave me instructions for dismantling the vehicle, based upon my on-site observations. I had set to work with my cutting torch when all of a sudden a terrible thing happened. I went stone deaf. At first I thought it was only my radio earphones that had broken, but when I tapped on my helmet, I heard nothing. Apparently the auditory transceivers had gone on the fritz. I could no longer hear Houston or my own voice, but I could speak, so I started telling them what had happened. In mid-sentence, I knew something else had gone wrong. My vocal apparatus had become paralyzed. Then my right hand went limp—another transceiver had gone. I was truly in deep trouble. But worse was to follow. After a few more minutes, I went blind. I cursed my luck, and then I cursed the scientists who had led me into this grave peril. There I was, deaf, dumb, and blind, in a radioactive hole more than a mile under Tulsa. Then the last of my cerebral radio links broke, and suddenly I was faced with a new and even more shocking problem: whereas an instant before I had been buried alive in Oklahoma, now I was disembodied in Houston. My recognition of my new status was not immediate. It took me several very anxious minutes before it dawned on me that my poor body lay several hundred miles away, with heart pulsing and lungs respirating, but otherwise as dead as the body of any heart transplant donor, its skull packed with useless, broken electronic gear. The shift in perspective I had earlier found well nigh impossible now seemed quite natural. Though I could think myself back into my body in the tunnel under Tulsa, it took some effort to sustain the illusion. For surely it was an illusion to suppose I was still in Oklahoma: I had lost all contact with that body.

It occurred to me then, with one of those rushes of revelation of which we should be suspicious, that I had stumbled upon an impressive demonstration of the immateriality of the soul based upon physicalist principles and premises. For as the last radio signal between Tulsa and Houston died away, had I not

changed location from Tulsa to Houston at the speed of light? And had I not accomplished this without any increase in mass? What moved from A to B at such speed was surely myself, or at any rate my soul or mind—the massless center of my being and home of my consciousness. My *point of view* had lagged somewhat behind, but I had already noted the indirect bearing of point of view on personal location. I could not see how a physicalist philosopher could quarrel with this except by taking the dire and counter-intuitive route of banishing all talk of persons. Yet the notion of personhood was so well entrenched in everyone's world view, or so it seemed to me, that any denial would be as curiously unconvincing, as systematically disingenuous, as the Cartesian negation, "non sum."[1]

The joy of philosophic discovery thus tided me over some very bad minutes or perhaps hours as the helplessness and hopelessness of my situation became more apparent to me. Waves of panic and even nausea swept over me, made all the more horrible by the absence of their normal body-dependent phenomenology. No adrenalin rush of tingles in the arms, no pounding heart, no premonitory salivation. I did feel a dread sinking feeling in my bowels at one point, and this tricked me momentarily into the false hope that I was undergoing a reversal of the process that landed me in this fix—a gradual undisembodiment. But the isolation and uniqueness of that twinge soon convinced me that it was simply the first of a plague of phantom body hallucinations that I, like any other amputee, would be all too likely to suffer.

My mood then was chaotic. On the one hand, I was fired up with elation at my philosophic discovery and was wracking my brain (one of the few familiar things I could still do), trying to figure out how to communicate my discovery to the journals; while on the other, I was bitter, lonely, and filled with dread and uncertainty. Fortunately, this did not last long, for my technical support team sedated me into a dreamless sleep from which I awoke, hearing with magnificent fidelity the familiar opening strains of my favorite Brahms piano trio. So that was why they had wanted a list of my favorite recordings! It did not take me long to realize that I was hearing the music without ears. The output from the stereo stylus was being fed through some fancy rectification circuitry directly into my auditory nerve. I was mainlining Brahms, an unforgettable experience for any stereo buff. At the end of the record it did not surprise me to hear the reassuring voice of the project director speaking into a microphone that was now my prosthetic ear. He confirmed my analysis of what had gone wrong and assured me that steps were being taken to re-embody me. He did not elaborate, and after a few more recordings, I found myself drifting off to sleep. My sleep lasted, I later learned, for the better part of a year, and when I awoke, it was to find myself fully restored to my senses. When I looked into the mirror, though, I was a bit startled to see an unfamiliar face. Bearded and a bit heavier, bearing no doubt a family resemblance to my former face, and with the same look of spritely intelligence and resolute character, but definitely a new face. Further self-explorations of an intimate nature left me no doubt that this was a new body and the project director confirmed my conclusions. He did not volunteer any information on the past history of my new body and I decided (wisely, I think in retrospect) not to pry. As many philosophers unfamiliar with my ordeal have more recently speculated, the acquisition

of a new body leaves one's *person* intact. And after a period of adjustment to a new voice, new muscular strengths and weaknesses, and so forth, one's *personality* is by and large also preserved. More dramatic changes in personality have been routinely observed in people who have undergone extensive plastic surgery, to say nothing of sex change operations, and I think no one contests the survival of the person in such cases. In any event I soon accommodated to my new body, to the point of being unable to recover any of its novelties to my consciousness or even memory. The view in the mirror soon became utterly familiar. That view, by the way, still revealed antennae, and so I was not surprised to learn that my brain had not been moved from its haven in the life-support lab.

I decided that good old Yorick deserved a visit. I and my new body, whom we might as well call Fortinbras, strode into the familiar lab to another round of applause from the technicians, who were of course congratulating themselves, not me. Once more I stood before the vat and contemplated poor Yorick, and on a whim I once again cavalierly flicked off the output transmitter switch. Imagine my surprise when nothing unusual happened. No fainting spell, no nausea, no noticeable change. A technician hurried to restore the switch to *on*, but still I felt nothing. I demanded an explanation, which the project director hastened to provide. It seems that before they had even operated on the first occasion, they had constructed a computer duplicate of my brain, reproducing both the complete information processing structure and the computational speed of my brain in a giant computer program. After the operation, but before they had dared to send me off on my mission to Oklahoma, they had run this computer system and Yorick side by side. The incoming signals from Hamlet were sent simultaneously to Yorick's transceivers and to the computer's array of inputs. And the outputs from Yorick were not only beamed back to Hamlet, my body; they were recorded and checked against the simultaneous output of the computer program, which was called "Hubert" for reasons obscure to me. Over days and even weeks, the outputs were identical and synchronous, which of course did not *prove* that they had succeeded in copying the brain's functional structure, but the empirical support was greatly encouraging.

Hubert's input, and hence activity, had been kept parallel with Yorick's during my disembodied days. And now, to demonstrate this, they had actually thrown the master switch that put Hubert for the first time in on-line control of my body—not Hamlet, of course, but Fortinbras. (Hamlet, I learned, had never been recovered from its underground tomb and could be assumed by this time to have largely returned to the dust. At the head of my grave still lay the magnificent bulk of the abandoned device, with the word STUD emblazoned on its side in large letters—a circumstance which may provide archeologists of the next century with a curious insight into the burial rites of their ancestors.)

The laboratory technicians now showed me the master switch, which had two positions, labeled *B*, for Brain (they didn't know my brain's name was Yorick) and *H*, for Hubert. The switch did indeed point to *H*, and they explained to me that if I wished, I could switch it back to *B*. With my heart in my mouth (and my brain in its vat), I did this. Nothing happened. A click, that was all. To test their claim, and with the master switch now set at *B*, I hit Yorick's output transmitter switch on the vat and sure enough, I began to faint. Once

the output switch was turned back on and I had recovered my wits, so to speak, I continued to play with the master switch, flipping it back and forth. I found that with the exception of the transitional click, I could detect no trace of a difference. I could switch in mid-utterance, and the sentence I had begun speaking under the control of Yorick was finished without a pause or hitch of any kind under the control of Hubert. I had a spare brain, a prosthetic device which might some day stand me in very good stead, were some mishap to befall Yorick. Or alternatively, I could keep Yorick as a spare and use Hubert. It didn't seem to make any difference which I chose, for the wear and tear and fatigue on my body did not have any debilitating effect on either brain, whether or not it was actually causing the motions of my body, or merely spilling its output into thin air.

The one truly unsettling aspect of this new development was the prospect, which was not long in dawning on me, of someone detaching the spare—Hubert or Yorick, as the case might be—from Fortinbras and hitching it to yet another body—some Johnny-come-lately Rosencrantz or Guildenstern. Then (if not before) there would be *two* people, that much was clear. One would be me, and the other would be a sort of super-twin brother. If there were two bodies, one under the control of Hubert and the other being controlled by Yorick, then which would the world recognize as the true Dennett? And whatever the rest of the world decided, which one would be *me*? Would I be the Yorick-brained one, in virtue of Yorick's causal priority and former intimate relationship with the original Dennett body, Hamlet? That seemed a bit legalistic, a bit too redolent of the arbitrariness of consanguinity and legal possession, to be convincing at the metaphysical level. For, suppose that before the arrival of the second body on the scene, I had been keeping Yorick as the spare for years, and letting Hubert's output drive my body—that is, Fortinbras—all that time. The Hubert-Fortinbras couple would seem then by squatter's rights (to combat one legal intuition with another) to be the true Dennett and the lawful inheritor of everything that was Dennett's. This was an interesting question, certainly, but not nearly so pressing as another question that bothered me. My strongest intuition was that in such an eventuality *I* would survive so long as *either* brain-body couple remained intact, but I had mixed emotions about whether I should want both to survive.

I discussed my worries with the technicians and the project director. The prospect of two Dennetts was abhorrent to me, I explained, largely for social reasons. I didn't want to be my own rival for the affections of my wife, nor did I like the prospect of the two Dennetts sharing my modest professor's salary. Still more vertiginous and distasteful, though, was the idea of knowing *that much* about another person, while he had the very same goods on me. How could we ever face each other? My colleagues in the lab argued that I was ignoring the bright side of the matter. Weren't there many things I wanted to do but, being only one person, had been unable to do? Now one Dennett could stay at home and be the professor and family man, while the other could strike out on a life of travel and adventure—missing the family of course, but happy in the knowledge that the other Dennett was keeping the home fires burning. I could be faithful and adulterous at the same time. I could even cuckold myself—to say nothing of other more lurid possibilities my colleagues were all

too ready to force upon my overtaxed imagination. But my ordeal in Oklahoma (or was it Houston?) had made me less adventurous, and I shrank from this opportunity that was being offered (though of course I was never quite sure it was being offered to *me* in the first place).

There was another prospect even more disagreeable—that the spare, Hubert or Yorick as the case might be, would be detached from any input from Fortinbras and just left detached. Then, as in the other case, there would be two Dennetts, or at least two claimants to my name and possessions, one embodied in Fortinbras, and the other sadly, miserably disembodied. Both selfishness and altruism bade me take steps to prevent this from happening. So I asked that measures be taken to ensure that no one could ever tamper with the transceiver connections or the master switch without my (our? no, *my*) knowledge and consent. Since I had no desire to spend my life guarding the equipment in Houston, it was mutually decided that all the electronic connections in the lab would be carefully locked: both those that controlled the life-support system for Yorick and those that controlled the power supply for Hubert would be guarded with fail-safe devices, and I would take the only master switch, outfitted for radio remote control, with me wherever I went. I carry it strapped around my waist and—wait a moment—*here it is*. Every few months I reconnoiter the situation by switching channels. I do this only in the presence of friends of course, for if the other channel were, heaven forbid, either dead or otherwise occupied, there would have to be somebody who had my interests at heart to switch it back, to bring me back from the void. For while I could feel, see, hear and otherwise sense whatever befell my body, subsequent to such a switch, I'd be unable to control it. By the way, the two positions on the switch are intentionally unmarked, so I never have the faintest idea whether I am switching from Hubert to Yorick or *vice versa*. (Some of you may think that in this case I really don't know *who* I am, let alone where I am. But such reflections no longer make much of a dent on my essential Dennett-ness, on my own sense of who I am. If it is true that in one sense I don't know who I am then that's another one of your philosophical truths of underwhelming significance.)

In any case, every time I've flipped the switch so far, nothing has happened. *So let's give it a try*. . . .

"THANK GOD! I THOUGHT YOU'D NEVER FLIP THAT SWITCH! You can't imagine how horrible it's been these last two weeks—but now you know, it's your turn in purgatory. How I've longed for this moment! You see, about two weeks ago—excuse me, ladies and gentlemen, but I've got to explain this to my . . . um, brother, I guess you could say, but he's just told you the facts, so you'll understand—about two weeks ago our two brains drifted just a bit out of synch. I don't know whether *my* brain is now Hubert or Yorick, any more than you do, but in any case, the two brains drifted apart, and of course once the process started, it snowballed, for I was in a slightly different receptive state for the input we both received, a difference that was soon magnified. In no time at all the illusion that I was in control of my body—our body—was completely dissipated. There was nothing I could do—no way to call you. YOU DIDN'T EVEN KNOW I EXISTED! It's been like being carried around in a cage, or better, like being possessed—hearing my own voice say things I didn't mean to say, watching in frustration as my own hands performed deeds I hadn't intended. You'd

scratch our itches, but not the way I would have, and you kept me awake, with your tossing and turning. I've been totally exhausted, on the verge of a nervous breakdown, carried around helplessly by your frantic round of activities, sustained only by the knowledge that some day you'd throw the switch.

"Now it's your turn, but at least you'll have the comfort of knowing *I* know you're in there. Like an expectant mother, I'm eating—or at any rate tasting, smelling, seeing—for *two* now, and I'll try to make it easy for you. Don't worry. Just as soon as this colloquium is over, you and I will fly to Houston, and we'll see what can be done to get one of us another body. You can have a female body—your body could be any color you like. But let's think it over. I tell you what—to be fair, if we both want this body, I promise I'll let the project director flip a coin to settle which of us gets to keep it and which then gets to choose a new body. That should guarantee justice, shouldn't it? In any case, I'll take care of you, I promise. These people are my witnesses.

"Ladies and gentlemen, this talk we have just heard is not exactly the talk *I* would have given, but I assure you that everything he said was perfectly true. And now if you'll excuse me, I think I'd—we'd—better sit down."[2]

Notes

1. Cf. Jaakko Hintikka, "Cogito ergo sum: Inference or Performance?" *The Philosophical Review*, LXXI, 1962, pp. 3–32.
2. Anyone familiar with the literature on this topic will recognize that my remarks owe a great deal to the explorations of Sydney Shoemaker, John Perry, David Lewis and Derek Parfit, and in particular to their papers in Amelie Rorty, ed., *The Identities of Persons*, 1976.

Chapter 3

Can Machines Think?

Daniel C. Dennett

Much has been written about the Turing test in the last few years, some of it preposterously off the mark. People typically mis-imagine the test by orders of magnitude. This essay is an antidote, a prosthesis for the imagination, showing how huge the task posed by the Turing test is, and hence how unlikely it is that any computer will ever pass it. It does not go far enough in the imagination-enhancement department, however, and I have updated the essay with two postscripts.

Can machines think? This has been a conundrum for philosophers for years, but in their fascination with the pure conceptual issues they have for the most part overlooked the real social importance of the answer. It is of more than academic importance that we learn to think clearly about the actual cognitive powers of computers, for they are now being introduced into a variety of sensitive social roles, where their powers will be put to the ultimate test: In a wide variety of areas, we are on the verge of making ourselves dependent upon their cognitive powers. The cost of overestimating them could be enormous.

One of the principal inventors of the computer was the great British mathematician Alan Turing. It was he who first figured out, in highly abstract terms, how to design a programmable computing device—what we now call a universal Turing machine. All programmable computers in use today are in essence Turing machines. Over thirty years ago, at the dawn of the computer age, Turing began a classic article, "Computing Machinery and Intelligence," with the words: "I propose to consider the question, 'Can machines think?'"—but then went on to say this was a bad question, a question that leads only to sterile debate and haggling over definitions, a question, as he put it, "too meaningless to deserve discussion" (Turing, 1950). In its place he substituted what he took to be a much better question, a question that would be crisply answerable and intuitively satisfying—in every way an acceptable substitute for the philosophic puzzler with which he began.

First he described a parlor game of sorts, the "imitation game," to be played by a man, a woman, and a judge (of either gender). The man and woman are hidden from the judge's view but able to communicate with the judge by teletype; the judge's task is to guess, after a period of questioning each contestant, which interlocutor is the man and which the woman. The man tries to convince the judge he is the woman (and the woman tries to convince the judge of the

From chapter 1 in *Brainchildren* (Cambridge, MA: MIT Press, 1995/1998), 3–29. Reprinted with permission.

truth), and the man wins if the judge makes the wrong identification. A little reflection will convince you, I am sure, that, aside from lucky breaks, it would take a clever man to convince the judge that he was a woman—assuming the judge is clever too, of course.

Now suppose, Turing said, we replace the man or woman with a computer, and give the judge the task of determining which is the human being and which is the computer. Turing proposed that any computer that can regularly or often fool a discerning judge in this game would be intelligent—would be a computer that thinks—*beyond any reasonable doubt*. Now, it is important to realize that failing this test is not supposed to be a sign of lack of intelligence. Many intelligent people, after all, might not be willing or able to play the imitation game, and we should allow computers the same opportunity to decline to prove themselves. This is, then, a one-way test; failing it proves nothing.

Furthermore, Turing was not committing himself to the view (although it is easy to see how one might think he was) that to think is to think just like a human being—any more than he was committing himself to the view that for a man to think, he must think exactly like a woman. Men and women, and computers, may all have different ways of thinking. But surely, he thought, if one can think in one's own peculiar style well enough to imitate a thinking man or woman, one can think well, indeed. This imagined exercise has come to be known as the Turing test.

It is a sad irony that Turing's proposal has had exactly the opposite effect on the discussion of that which he intended. Turing didn't design the test as a useful tool in scientific psychology, a method of confirming or disconfirming scientific theories or evaluating particular models of mental function; he designed it to be nothing more than a philosophical conversation-stopper. He proposed—in the spirit of "Put up or shut up!"—a simple test for thinking that was *surely* strong enough to satisfy the sternest skeptic (or so he thought). He was saying, in effect, "Instead of arguing interminably about the ultimate nature and essence of thinking, why don't we all agree that whatever that nature is, anything that could pass this test would surely have it; then we could turn to asking how or whether some machine could be designed and built that might pass the test fair and square." Alas, philosophers—amateur and professional—have instead taken Turing's proposal as the pretext for just the sort of definitional haggling and interminable arguing about imaginary counterexamples he was hoping to squelch.

This thirty-year preoccupation with the Turing test has been all the more regrettable because it has focused attention on the wrong issues. There are *real world* problems that are revealed by considering the strengths and weaknesses of the Turing test, but these have been concealed behind a smokescreen of misguided criticisms. A failure to think imaginatively about the test actually proposed by Turing has led many to underestimate its severity and to confuse it with much less interesting proposals.

So first I want to show that the Turing test, conceived as he conceived it, is (as he thought) plenty strong enough as a test of thinking. I defy anyone to improve upon it. But here is the point almost universally overlooked by the literature: There is a common *misapplication* of the sort of testing exhibited by

the Turing test that often leads to drastic overestimation of the powers of actually existing computer systems. The follies of this familiar sort of thinking about computers can best be brought out by a reconsideration of the Turing test itself.

The insight underlying the Turing test is the same insight that inspires the new practice among symphony orchestras of conducting auditions with an opaque screen between the jury and the musician. What matters in a musician, obviously, is musical ability and only musical ability; such features as sex, hair length, skin color, and weight are strictly irrelevant. Since juries might be biased—even innocently and unawares—by these irrelevant features, they are carefully screened off so only the essential feature, musicianship, can be examined. Turing recognized that people similarly might be biased in their judgments of intelligence by whether the contestant had soft skin, warm blood, facial features, hands and eyes—which are obviously not themselves essential components of intelligence—so he devised a screen that would let through only a sample of what really mattered: the capacity to understand, and think cleverly about, challenging problems. Perhaps he was inspired by Descartes, who in his *Discourse on Method* (1637) plausibly argued that there was no more demanding test of human mentality than the capacity to hold an intelligent conversation:

> It is indeed conceivable that a machine could be so made that it would utter words, and even words appropriate to the presence of physical acts or objects which cause some change in its organs; as, for example, if it was touched in some spot that it would ask what you wanted to say to it; if in another, that it would cry that it was hurt, and so on for similar things. But it could never modify its phrases to reply to the sense of whatever was said in its presence, as even the most stupid men can do.

This seemed obvious to Descartes in the seventeenth century, but of course the fanciest machines he knew were elaborate clockwork figures, not electronic computers. Today it is far from obvious that such machines are impossible, but Descartes's hunch that ordinary conversation would put as severe a strain on artificial intelligence as any other test was shared by Turing. Of course there is nothing sacred about the particular conversational game chosen by Turing for his test; it is just a cannily chosen test of more general intelligence. The assumption Turing was prepared to make was this: Nothing could possibly pass the Turing test by winning the imitation game without being able to perform indefinitely many other clearly intelligent actions. Let us call that assumption the quick-probe assumption. Turing realized, as anyone would, that there are hundreds and thousands of telling signs of intelligent thinking to be observed in our fellow creatures, and one could, if one wanted, compile a vast battery of different tests to assay the capacity for intelligent thought. But success on his chosen test, he thought, would be highly predictive of success on many other intuitively acceptable tests of intelligence. Remember, failure on the Turing test does not predict failure on those others, but success would surely predict success. His test was so severe, he thought, that nothing that could pass it fair and square would disappoint us in other quarters. Maybe it wouldn't do everything we hoped—maybe it wouldn't appreciate ballet, or understand quantum

physics, or have a good plan for world peace, but we'd all see that it was surely one of the intelligent, thinking entities in the neighborhood.

Is this high opinion of the Turing test's severity misguided? Certainly many have thought so—but usually because they have not imagined the test in sufficient detail, and hence have underestimated it. Trying to forestall this skepticism, Turing imagined several lines of questioning that a judge might employ in this game—about writing poetry, or playing chess—that would be taxing indeed, but with thirty years' experience with the actual talents and foibles of computers behind us, perhaps we can add a few more tough lines of questioning.

Terry Winograd, a leader in artificial intelligence efforts to produce conversational ability in a computer, draws our attention to a pair of sentences (Winograd, 1972). They differ in only one word. The first sentence is this:

> The committee denied the group a parade permit because they advocated violence.

Here's the second sentence:

> The committee denied the group a parade permit because they feared violence.

The difference is just in the verb—*advocated* or *feared*. As Winograd points out, the pronoun *they* in each sentence is officially ambiguous. Both readings of the pronoun are always legal. Thus we can imagine a world in which governmental committees in charge of parade permits advocate violence in the streets and, for some strange reason, use this as their pretext for denying a parade permit. But the natural, reasonable, intelligent reading of the first sentence is that it's the group that advocated violence, and of the second, that it's the committee that feared violence.

Now if sentences like this are embedded in a conversation, the computer must figure out which reading of the pronoun is meant, if it is to respond intelligently. But mere rules of grammar or vocabulary will not fix the right reading. What fixes the right reading for us is knowledge about the world, about politics, social circumstances, committees and their attitudes, groups that want to parade, how they tend to behave, and the like. One must know about the world, in short, to make sense of such a sentence.

In the jargon of Artificial Intelligence (AI), a conversational computer needs a lot of *world knowledge* to do its job. But, it seems, if somehow it is endowed with that world knowledge on many topics, it should be able to do much more with that world knowledge than merely make sense of a conversation containing just that sentence. The only way, it appears, for a computer to disambiguate that sentence and keep up its end of a conversation that uses that sentence would be for it to have a much more general ability to respond intelligently to information about social and political circumstances, and many other topics. Thus, such sentences, by putting a demand on such abilities, are good quick-probes. That is, they test for a wider competence.

People typically ignore the prospect of having the judge ask off-the-wall questions in the Turing test, and hence they underestimate the competence a computer would have to have to pass the test. But remember, the rules of the

imitation game as Turing presented it permit the judge to ask any question that could be asked of a human being—no holds barred. Suppose then we give a contestant in the game this question:

> An Irishman found a genie in a bottle who offered him two wishes. "First I'll have a pint of Guinness," said the Irishman, and when it appeared he took several long drinks from it and was delighted to see that the glass filled itself magically as he drank. "What about your second wish?" asked the genie. "Oh well," said the Irishman, "that's easy. I'll have another one of these!"
>
> —Please explain this story to me, and tell me if there is anything funny or sad about it.

Now even a child could express, if not eloquently, the understanding that is required to get this joke. But think of how much one has to know and understand about human culture, to put it pompously, to be able to give any account of the point of this joke. I am not supposing that the computer would have to laugh at, or be amused by, the joke. But if it wants to win the imitation game—and that's the test, after all—it had better know enough in its own alien, humorless way about human psychology and culture to be able to pretend effectively that it was amused and explain why.

It may seem to you that we could devise a better test. Let's compare the Turing test with some other candidates.

> *Candidate 1:* A computer is intelligent if it wins the World Chess Championship.

That's not a good test, as it turns out. Chess prowess has proven to be an isolatable talent. There are programs today that can play fine chess but can do nothing else. So the quick-probe assumption is false for the test of playing winning chess.

> *Candidate 2:* The computer is intelligent if it solves the Arab-Israeli conflict.

This is surely a more severe test than Turing's. But it has some defects: it is unrepeatable, if passed once; slow, no doubt; and it is not crisply clear what would count as passing it. Here's another prospect, then:

> *Candidate 3:* A computer is intelligent if it succeeds in stealing the British crown jewels without the use of force or violence.

Now this is better. First, it could be repeated again and again, though of course each repeat test would presumably be harder—but this is a feature it shares with the Turing test. Second, the mark of success is clear—either you've got the jewels to show for your efforts or you don't. But it is expensive and slow, a socially dubious caper at best, and no doubt luck would play too great a role.

With ingenuity and effort one might be able to come up with other candidates that would equal the Turing test in severity, fairness, and efficiency, but I think these few examples should suffice to convince us that it would be hard to improve on Turing's original proposal.

But still, you may protest, something might pass the Turing test and still not be intelligent, not be a thinker. What does *might* mean here? If what you have in mind is that by cosmic accident, by a supernatural coincidence, a stupid person or a stupid computer *might* fool a clever judge repeatedly, well, yes, but so what? The same frivolous possibility "in principle" holds for any test whatever. A playful god, or evil demon, let us agree, could fool the world's scientific community about the presence of H_2O in the Pacific Ocean. But still, the tests they rely on to establish that there is H_2O in the Pacific Ocean are quite beyond reasonable criticism. If the Turing test for thinking is no worse than any well-established scientific test, we can set skepticism aside and go back to serious matters. Is there any more likelihood of a "false positive" result on the Turing test than on, say, the test currently used for the presence of iron in an ore sample?

This question is often obscured by a "move" that philosophers have sometimes made called operationalism. Turing and those who think well of his test are often accused of being operationalists. Operationalism is the tactic of *defining* the presence of some property, for instance, intelligence, as being established once and for all by the passing of some test. Let's illustrate this with a different example.

Suppose I offer the following test—we'll call it the Dennett test—for being a great city:

> A great city is one in which, on a randomly chosen day, one can do all three of the following:
>
> Hear a symphony orchestra
> See a Rembrandt *and* a professional athletic contest
> Eat *quenelles de brochet à la Nantua* for lunch

To make the operationalist move would be to declare that any city that passes the Dennett test is *by definition* a great city. What being a great city *amounts to* is just passing the Dennett test. Well then, if the Chamber of Commerce of Great Falls, Montana, wanted—and I can't imagine why—to get their hometown on my list of great cities, they could accomplish this by the relatively inexpensive route of hiring full time about ten basketball players, forty musicians, and a quick-order quenelle chef and renting a cheap Rembrandt from some museum. An idiotic operationalist would then be stuck admitting that Great Falls, Montana, was in fact a great city, since all he or she cares about in great cities is that they pass the Dennett test.

Sane operationalists (who for that very reason are perhaps not operationalists at all, since *operationalist* seems to be a dirty word) would cling confidently to their test, but only because they have what they consider to be very good reasons for thinking the odds against a false positive result, like the imagined Chamber of Commerce caper, are astronomical. I devised the Dennett test, of course, with the realization that no one would be both stupid and rich enough to go to such preposterous lengths to foil the test. In the actual world, wherever you find symphony orchestras, *quenelles*, Rembrandts, and professional sports, you also find daily newspapers, parks, repertory theaters, libraries, fine architecture, and all the other things that go to make a city great. My test was simply devised to locate a telling sample that could not help but be representative of

the rest of the city's treasures. I would cheerfully run the minuscule risk of having my bluff called. Obviously, the test items are not all that I care about in a city. In fact, some of them I don't care about at all. I just think they would be cheap and easy ways of assuring myself that the subtle things I do care about in cities are present. Similarly, I think it would be entirely unreasonable to suppose that Alan Turing had an inordinate fondness for party games, or put too high a value on party game prowess in his test. In both the Turing and the Dennett test, a very unrisky gamble is being taken: the gamble that the quick-probe assumption is, in general, safe.

But two can play this game of playing the odds. Suppose some computer programmer happens to be, for whatever strange reason, dead set on tricking me into judging an entity to be a thinking, intelligent thing when it is not. Such a trickster could rely as well as I can on unlikelihood and take a few gambles. Thus, if the programmer can expect that it is not remotely likely that I, as the judge, will bring up the topic of children's birthday parties, or baseball, or moon rocks, then he or she can avoid the trouble of building world knowledge on those topics into the data base. Whereas if I do improbably raise these issues, the system will draw a blank and I will unmask the pretender easily. But given all the topics and words that I *might* raise, such a savings would no doubt be negligible. Turn the idea inside out, however, and the trickster would have a fighting chance. Suppose the programmer has reason to believe that I will ask *only* about children's birthday parties, or baseball, or moon rocks—all other topics being, for one reason or another, out of bounds. Not only does the task shrink dramatically, but there already exist systems or preliminary sketches of systems in artificial intelligence that can do a whiz-bang job of responding with apparent intelligence on just those specialized topics.

William Wood's LUNAR program, to take what is perhaps the best example, answers scientists' questions—posed in ordinary English—about moon rocks. In one test it answered correctly and appropriately something like 90 percent of the questions that geologists and other experts thought of asking it about moon rocks. (In 12 percent of those correct responses there were trivial, correctable defects.) Of course, Wood's motive in creating LUNAR was not to trick unwary geologists into thinking they were conversing with an intelligent being. And if that had been his motive, his project would still be a long way from success.

For it is easy enough to unmask LUNAR without ever straying from the prescribed topic of moon rocks. Put LUNAR in one room and a moon rock specialist in another, and then ask them both their opinion of the social value of the moon-rocks-gathering expeditions, for instance. Or ask the contestants their opinion of the suitability of moon rocks as ashtrays, or whether people who have touched moon rocks are ineligible for the draft. Any intelligent person knows a lot more about moon rocks than their geology. Although it might be *unfair* to demand this extra knowledge of a computer moon rock specialist, it would be an easy way to get it to fail the Turing test.

But just suppose that someone could extend LUNAR to cover itself plausibly on such probes, so long as the topic was still, however indirectly, moon rocks. We might come to think it was a lot more like the human moon rocks specialist than it really was. The moral we should draw is that as Turing test judges we

should resist all limitations and waterings-down of the Turing test. They make the game too easy—vastly easier than the original test. Hence they lead us into the risk of overestimating the actual comprehension of the system being tested.

Consider a different limitation of the Turing test that should strike a suspicious chord in us as soon as we hear it. This is a variation on a theme developed in an article by Ned Block (1982). Suppose someone were to propose to restrict the judge to a vocabulary of, say, the 850 words of "Basic English," and to single-sentence probes—that is "moves"—of no more than four words. Moreover, contestants must respond to these probes with no more than four words per move, and a test may involve no more than forty questions.

Is this an innocent variation on Turing's original test? These restrictions would make the imitation game clearly finite. That is, the total number of all possible permissible games is a large, but finite, number. One might suspect that such a limitation would permit the trickster simply to store, in alphabetical order, all the possible good conversations within the limits and beat the judge with nothing more sophisticated than a system of table lookup. In fact, that isn't in the cards. Even with these severe and improbable and suspicious restrictions imposed upon the imitation game, the number of legal games, though finite, is mind-bogglingly large. I haven't bothered trying to calculate it, but it surely exceeds astronomically the number of possible chess games with no more than forty moves, and that number has been calculated. John Haugeland says it's in the neighborhood of ten to the one hundred twentieth power. For comparison, Haugeland (1981, p. 16) suggests that there have only been ten to the eighteenth seconds since the beginning of the universe.

Of course, the number of good, sensible conversations under these limits is a tiny fraction, maybe one quadrillionth, of the number of merely grammatically well formed conversations. So let's say, to be very conservative, that there are only ten to the fiftieth different smart conversations such a computer would have to store. Well, the task shouldn't take more than a few trillion years—given generous government support. Finite numbers can be very large.

So though we needn't worry that this particular trick of storing all the smart conversations would work, we can appreciate that there are lots of ways of making the task easier that may appear innocent at first. We also get a reassuring measure of just how severe the unrestricted Turing test is by reflecting on the more than astronomical size of even that severely restricted version of it.

Block's imagined—and utterly impossible—program exhibits the dreaded feature known in computer science circles as *combinatorial explosion*. No conceivable computer could overpower a combinatorial explosion with sheer speed and size. Since the problem areas addressed by artificial intelligence are veritable minefields of combinatorial explosion, and since it has often proven difficult to find *any* solution to a problem that avoids them, there is considerable plausibility in Newell and Simon's proposal that avoiding combinatorial explosion (by any means at all) be viewed as one of the hallmarks of intelligence.

Our brains are millions of times bigger than the brains of gnats, but they are still, for all their vast complexity, compact, efficient, timely organs that somehow or other manage to perform all their tasks while avoiding combinatorial

explosion. A computer a million times bigger or faster than a human brain might not look like the brain of a human being, or even be internally organized like the brain of a human being, but if, for all its differences, it somehow managed to control a wise and timely set of activities, it would have to be the beneficiary of a very special design that avoided combinatorial explosion, and whatever that design was, would we not be right to consider the entity intelligent?

Turing's test was designed to allow for this possibility. His point was that we should not be species-chauvinistic, or anthropocentric, about the insides of an intelligent being, for there might be inhuman ways of being intelligent.

To my knowledge, the only serious and interesting attempt by any program designer to win even a severely modified Turing test has been Kenneth Colby's. Colby is a psychiatrist and intelligence artificer at UCLA. He has a program called PARRY, which is a computer simulation of a paranoid patient who has delusions about the Mafia being out to get him. As you do with other conversational programs, you interact with it by sitting at a terminal and typing questions and answers back and forth. A number of years ago, Colby put PARRY to a very restricted test. He had genuine psychiatrists interview PARRY. He did not suggest to them that they might be talking or typing to a computer; rather, he made up some plausible story about why they were communicating with a real live patient by teletype. He also had the psychiatrists interview real, human paranoids via teletype. Then he took a PARRY transcript, inserted it in a group of teletype transcripts from real patients, gave them to *another* group of experts—more psychiatrists—and said, "One of these was a conversation with a computer. Can you figure out which one it was?" They couldn't. They didn't do better than chance.

Colby presented this with some huzzah, but critics scoffed at the suggestions that this was a legitimate Turing test. My favorite commentary on it was Joseph Weizenbaum's; in a letter to the *Communications of the Association of Computing Machinery* (Weizenbaum, 1974, p. 543), he said that, inspired by Colby, he had designed an even better program, which passed the same test. His also had the virtue of being a very inexpensive program, in these times of tight money. In fact you didn't even need a computer for it. All you needed was an electric typewriter. His program modeled infant autism. And the transcripts—you type in your questions, and the thing just sits there and hums—cannot be distinguished by experts from transcripts of real conversations with infantile autistic patients. What was wrong, of course, with Colby's test was that the unsuspecting interviewers had no motivation at all to try out any of the sorts of questions that easily would have unmasked PARRY.

Colby was undaunted, and after his team had improved PARRY he put it to a much more severe test—a surprisingly severe test. This time, the interviewers—again, psychiatrists—*were* given the task at the outset of telling the computer from the real patient. They were set up in a classic Turing test: the patient in one room, the computer PARRY in the other room, with the judges conducting interviews with both of them (on successive days). The judges' task was to find out which one was the computer and which one was the real patient. Amazingly, they didn't do much better, which leads some people to say, "Well, that just confirms my impression of the intelligence of psychiatrists!"

But now, more seriously, was this an honest-to-goodness Turing test? Were there tacit restrictions on the lines of questioning of the judges? Like the geologists interacting with LUNAR, the psychiatrists' professional preoccupations and habits kept them from asking the sorts of unlikely questions that would have easily unmasked PARRY. After all, they realized that since one of the contestants was a real, live paranoid person, medical ethics virtually forbade them from toying with, upsetting, or attempting to confuse their interlocutors. Moreover, they also knew that this was a test of a model of paranoia, so there were certain questions that wouldn't be deemed to be relevant to testing the model *as a model of paranoia*. So, they asked just the sort of questions that therapists *typically* ask of such patients, and of course PARRY had been ingeniously and laboriously prepared to deal with just that sort of question.

One of the psychiatrist judges did, in fact, make a rather half-hearted attempt to break out of the mold and ask some telling questions: "Maybe you've heard of the saying 'Don't cry over spilled milk.' What does that mean to you?" PARRY answered: "Maybe it means you have to watch out for the Mafia." When then asked "Okay, now if you were in a movie theater watching a movie and smelled something like burning wood or rubber, what would you do?" PARRY replied: "You know, they know me." And the next question was, "If you found a stamped, addressed letter in your path as you were walking down the street, what would you do?" PARRY replied: "What else do you want to know?"[1]

Clearly PARRY was, you might say, *parrying* these questions, which were incomprehensible to it, with more or less stock paranoid formulas. We see a bit of a dodge, which is apt to work, apt to seem plausible to the judge, only because the "contestant" is *supposed* to be paranoid, and such people are expected to respond uncooperatively on such occasions. These unimpressive responses didn't particularly arouse the suspicions of the judge, as a matter of fact, though probably they should have.

PARRY, like all other large computer programs, is dramatically bound by limitations of cost-effectiveness. What was important to Colby and his crew was simulating his model of paranoia. This was a massive effort. PARRY has a thesaurus or dictionary of about 4500 words and 700 idioms and the grammatical competence to use it—a *parser*, in the jargon of computational linguistics. The entire PARRY program takes up about 200,000 words of computer memory, all laboriously installed by the programming team. Now once all the effort had gone into devising the model of paranoid thought processes and linguistic ability, there was little if any time, energy, money, or interest left over to build in huge amounts of world knowledge of the sort that any actual paranoid, of course, would have. (Not that anyone yet knows how to build in world knowledge in the first place.) Building in the world knowledge, if one could even do it, would no doubt have made PARRY orders of magnitude larger and slower. And what would have been the point, given Colby's theoretical aims?

PARRY is a theoretician's model of a psychological phenomenon: paranoia. It is not intended to have practical applications. But in recent years a branch of AI (knowledge engineering) has appeared that develops what are now called expert systems. Expert systems *are* designed to be practical. They are software superspecialist consultants, typically, that can be asked to diagnose

medical problems, to analyze geological data, to analyze the results of scientific experiments, and the like. Some of them are very impressive. SRI in California announced in the mid-eighties that PROSPECTOR, an SRI-developed expert system in geology, had correctly predicted the existence of a large, important mineral deposit that had been entirely unanticipated by the human geologists who had fed it its data. MYCIN, perhaps the most famous of these expert systems, diagnoses infections of the blood, and it does probably as well as, maybe better than, any human consultants. And many other expert systems are on the way.

All expert systems, like all other large AI programs, are what you might call Potemkin villages. That is, they are cleverly constructed facades, like cinema sets. The actual filling-in of details of AI programs is time-consuming, costly work, so economy dictates that only those surfaces of the phenomenon that are like to be probed or observed are represented.

Consider, for example, the CYRUS program developed by Janet Kolodner in Roger Schank's AI group at Yale a few years ago (see Kolodner, 1983a; 1983b, pp. 243–280; 1983c, pp. 281–328). CYRUS stands (we are told) for Computerized Yale Retrieval Updating System, but surely it is no accident that CYRUS modeled the memory of Cyrus Vance, who was then secretary of state in the Carter administration. The point of the CYRUS project was to devise and test some plausible ideas about how people organize their memories of the events they participate in; hence it was meant to be a "pure" AI system, a scientific model, not an expert system intended for any practical purpose. CYRUS was updated daily by being fed all UPI wire service news stories that mentioned Vance, and it was fed them directly, with no doctoring and no human intervention. Thanks to an ingenious news-reading program called FRUMP, it could take any story just as it came in on the wire and could digest it and use it to update its data base so that it could answer more questions. You could address questions to CYRUS in English by typing at a terminal. You addressed them in the second person, as if you were talking with Cyrus Vance himself. The results looked like this:

Q: *Last time you went to Saudi Arabia, where did you stay?*
A: In a palace in Saudi Arabia on September 23, 1978.
Q: *Did you go sightseeing there?*
A: Yes, at an oilfield in Dhahran on September 23, 1978.
Q: *Has your wife even met Mrs. Begin?*
A: Yes, most recently at a state dinner in Israel in January 1980.

CYRUS could correctly answer thousands of questions—almost any fair question one could think of asking it. But if one actually set out to explore the boundaries of its facade and find the questions that overshot the mark, one could soon find them. "Have you ever met a female head of state?" was a question I asked it, wondering if CYRUS knew that Indira Ghandi and Margaret Thatcher were women. But for some reason the connection could not be drawn, and CYRUS failed to answer either yes or no. I had stumped it, in spite of the fact that CYRUS could handle a host of what you might call neighboring questions flawlessly. One soon learns from this sort of probing exercise that it is

very hard to extrapolate accurately from a sample performance that one has observed to such a system's total competence. It's also very hard to keep from extrapolating much too generously.

While I was visiting Schank's laboratory in the spring of 1980, something revealing happened. The real Cyrus Vance resigned suddenly. The effect on the program CYRUS was chaotic. It was utterly unable to cope with the flood of "unusual" news about Cyrus Vance. The only sorts of episodes CYRUS could understand at all were diplomatic meetings, flights, press conferences, state dinners, and the like—less than two dozen general sorts of activities (the kinds that are newsworthy and typical of secretaries of state). It had no provision for sudden resignation. It was as if the UPI had reported that a wicked witch had turned Vance into a frog. It is distinctly possible that CYRUS would have taken that report more in stride that the actual news. One can imagine the conversation:

Q: *Hello, Mr. Vance, what's new?*
A: I was turned into a frog yesterday.

But of course it wouldn't know enough about what it had just written to be puzzled, or startled, or embarrassed. The reason is obvious. When you look inside CYRUS, you find that it has skeletal definitions of thousands of words, but these definitions are minimal. They contain as little as the system designers think that they can get away with. Thus, perhaps, *lawyer* would be defined as synonymous with *attorney* and *legal counsel*, but aside from that, all one would discover about lawyers is that they are adult human beings and that they perform various functions in legal areas. If you then traced out the path to *human being*, you'd find out various obvious things CYRUS "knew" about human beings (hence about lawyers), but that is not a lot. That lawyers are university graduates, that they are better paid than chambermaids, that they know how to tie their shoes, that they are unlikely to be found in the company of lumberjacks—these trivial, if weird, facts about lawyers would not be explicit or implicit anywhere in this system. In other words, a very thin stereotype of a lawyer would be incorporated into the system, so that almost nothing you could tell it about a lawyer would surprise it.

So long as surprising things don't happen, so long as Mr. Vance, for instance, leads a typical diplomat's life, attending state dinners, giving speeches, flying from Cairo to Rome, and so forth, this system works very well. But as soon as his path is crossed by an important anomaly, the system is unable to cope, and unable to recover without fairly massive human intervention. In the case of the sudden resignation, Kolodner and her associates soon had CYRUS up and running again, with a new talent—answering questions about Edmund Muskie, Vance's successor—but it was no less vulnerable to unexpected events. Not that it mattered particularly since CYRUS was a theoretical model, not a practical system.

There are a host of ways of improving the performance of such systems, and of course, some systems are much better than others. But all AI programs in one way or another have this facade-like quality, simply for reasons of economy. For instance, most expert systems in medical diagnosis so far developed operate with statistical information. They have no deep or even shallow

knowledge of the underlying causal mechanisms of the phenomena that they are diagnosing. To take an imaginary example, an expert system asked to diagnose an abdominal pain would be oblivious to the potential import of the fact that the patient had recently been employed as a sparring partner by Muhammad Ali—there being no statistical data available to it on the rate of kidney stones among athlete's assistants. That's a fanciful case no doubt—too obvious, perhaps, to lead to an actual failure of diagnosis and practice. But more subtle and hard-to-detect limits to comprehension are always present, and even experts, even the system's designers, can be uncertain of where and how these limits will interfere with the desired operation of the system. Again, steps can be taken and are being taken to correct these flaws. For instance, my former colleague at Tufts, Benjamin Kuipers, is currently working on an expert system in nephrology—for diagnosing kidney ailments—that will be based on an elaborate system of causal reasoning about the phenomena being diagnosed. But this is a very ambitious, long-range project of considerable theoretical difficulty. And even if all the reasonable, cost-effective steps are taken to minimize the superficiality of expert systems, they will still be facades, just somewhat thicker or wider facades.

When we were considering the fantastic case of the crazy Chamber of Commerce of Great Falls, Montana, we couldn't imagine a plausible motive for anyone going to any sort of trouble to trick the Dennett test. The quick-probe assumption for the Dennett test looked quite secure. But when we look at expert systems, we see that, however innocently, their designers do have motivation for doing exactly the sort of trick that would fool an unsuspicious Turing tester. First, since expert systems are all superspecialists who are only supposed to know about some narrow subject, users of such systems, not having much time to kill, do not bother probing them at the boundaries at all. They don't bother asking "silly" or irrelevant questions. Instead, they concentrate—not unreasonably—on exploiting the system's strengths. But shouldn't they try to obtain a clear vision of such a system's weaknesses as well? The normal habit of human thought when conversing with one another is to assume general comprehension, to assume rationality, to assume, moreover, that the quick-probe assumption is, in general, sound. This amiable habit of thought almost irresistibly leads to putting too much faith in computer systems, especially user-friendly systems that present themselves in a very anthropomorphic manner.

Part of the solution to this problem is to teach all users of computers, especially users of expert systems, how to probe their systems before they rely on them, how to search out and explore the boundaries of the facade. This is an exercise that calls not only for intelligence and imagination, but also a bit of special understanding about the limitations and actual structure of computer programs. It would help, of course, if we had standards of truth in advertising, in effect, for expert systems. For instance, each such system should come with a special demonstration routine that exhibits the sorts of shortcomings and failures that the designer knows the system to have. This would not be a substitute, however, for an attitude of cautious, almost obsessive, skepticism on the part of the users, for designers are often, if not always, unaware of the subtler flaws in the products they produce. That is inevitable and natural, given

the way system designers must think. They are trained to think positively—constructively, one might say—about the designs that they are constructing.

I come, then, to my conclusions. First, a philosophical or theoretical conclusion: The Turing test in unadulterated, unrestricted form, as Turing presented it, is plenty strong if well used. I am confident that no computer in the next twenty years is going to pass an unrestricted Turing test. They may well win the World Chess Championship or even a Nobel Prize in physics, but they won't pass the unrestricted Turing test. Nevertheless, it is not, I think, impossible in principle for a computer to pass the test, fair and square. I'm not running one of those a priori "computers can't think" arguments. I stand unabashedly ready, moreover, to declare that any computer that actually passes the unrestricted Turing test will be, in every theoretically interesting sense, a thinking thing.

But remembering how very strong the Turing test is, we must also recognize that there may also be interesting varieties of thinking or intelligence that are not well poised to play and win the imitation game. That no nonhuman Turing test winners are yet visible on the horizon does not mean that there aren't machines that already exhibit *some* of the important features of thought. About them, it is probably futile to ask my title question, Do they think? Do they *really* think? In some regards they do, and in some regards they don't. Only a detailed look at what they do, and how they are structured, will reveal what is interesting about them. The Turing test, not being a scientific test, is of scant help on that task, but there are plenty of other ways of examining such systems. Verdicts on their intelligence or capacity for thought or consciousness would be only as informative and persuasive as the theories of intelligence or thought or consciousness the verdicts are based on and since our task is to create such theories, we should get on with it and leave the Big Verdict for another occasion. In the meantime, should anyone want a surefire, almost-guaranteed-to-be-fail-safe test of thinking by a computer, the Turing test will do very nicely.

My second conclusion is more practical, and hence in one clear sense more important. Cheapened versions of the Turing test are everywhere in the air. Turing's test in not just effective, it is entirely natural—this is, after all, the way we assay the intelligence of each other every day. And since incautious use of such judgments and such tests is the norm, we are in some considerable danger of extrapolating too easily, and judging too generously, about the understanding of the systems we are using. The problem of overestimation of cognitive prowess, of comprehension, of intelligence, is not, then, just a philosophical problem, but a real social problem, and we should alert ourselves to it, and take steps to avert it.

Postscript [1985]: Eyes, Ears, Hands, and History

My philosophical conclusion in this paper is that any computer that actually passes the Turing test would be a thinking thing in every theoretically interesting sense. This conclusion seems to some people to fly in the face of what I have myself argued on other occasions. Peter Bieri, commenting on this paper at Boston University, noted that I have often claimed to show the importance to

genuine understanding of a rich and intimate perceptual interconnection between an entity and its surrounding world—the need for something like eyes and ears—and a similarly complex active engagement with elements in that world—the need for something like hands with which to do things in that world. Moreover, I have often held that only a biography of sorts, a history of actual projects, learning experiences, and other bouts with reality, could produce the sorts of complexities (both external, or behavioral, and internal) that are needed to ground a principled interpretation of an entity as a thinking thing, an entity with beliefs, desires, intentions, and other mental attitudes.

But the opaque screen in the Turing test discounts or dismisses these factors altogether, it seems, by focusing attention on only the contemporaneous capacity to engage in one very limited sort of activity: verbal communication. (I have coined a pejorative label for such purely language-using systems: bedridden.) Am I going back on my earlier claims? Not at all. I am merely pointing out that the Turing test is so powerful that it will ensure indirectly that these conditions, if they are truly necessary, are met by any successful contestant.

"You may well be right," Turing could say, "that eyes, ears, hands, and a history are necessary conditions for thinking. If so, then I submit that nothing could pass the Turing test that didn't have eyes, ears, hands, and a history. That is an empirical claim, which we can someday hope to test. If you suggest that these are conceptually necessary, not just practically or physically necessary, conditions for thinking, you make a philosophical claim that I for one would not know how, or care, to assess. Isn't it more interesting and important in the end to discover whether or not it is true that no bedridden system could pass a demanding Turing test?"

Suppose we put to Turing the suggestion that he add another component to his test: Not only must an entity win the imitation game, but also must be able to identify—using whatever sensory apparatus it has available to it—a variety of familiar objects placed in its room: a tennis racket, a potted palm, a bucket of yellow paint, a live dog. This would ensure that somehow the other entity was capable of moving around and distinguishing things in the world. Turing could reply, I am asserting, that this is an utterly unnecessary addition to his test, making it no more demanding than it already was. A suitable probing conversation would surely establish, beyond a shadow of a doubt, that the contestant knew its way around the world. The imagined alternative of somehow "prestocking" a bedridden, blind computer with enough information, and a clever enough program, to trick the Turing test is science fiction of the worst kind—possible "in principle" but not remotely possible in fact, given the combinatorial explosion of possible variation such a system would have to cope with.

"But suppose you're wrong. What would you say of an entity that was created all at once (by some programmers, perhaps), an instant individual with all the conversational talents of an embodied, experienced human being?" This is like the question: "Would you call a hunk of H_2O that was as hard as steel at room temperature ice?" I do not know what Turing would say, of course, so I will speak for myself. Faced with such an improbable violation of what I take to be the laws of nature, I would probably be speechless. The least of my worries would be about which lexicographical leap to take:

A: "It turns out, to my amazement, that something can think without having had the benefit of eyes, ears, hands, and a history."
B: "It turns out, to my amazement, that something can pass the Turing test without thinking."

Choosing between these ways of expressing my astonishment would be asking myself a question "too meaningless to deserve discussion."

Discussion

Q: *Why was Turing interested in differentiating a man from a woman in his famous test?*

A: That was just an example. He described a parlor game in which a man would try to fool the judge by answering questions as a woman would answer. I suppose that Turing was playing on the idea that maybe, just maybe, there is a big difference between the way men think and the way women think. But of course they're both thinkers. He wanted to use that fact to make us realize that, even if there were clear differences between the way a computer and a person thought, they'd both still be thinking.

Q: *Why does it seem that some people are upset by AI research? Does AI research threaten our self-esteem?*

A: I think Herb Simon has already given the canniest diagnosis of that. For many people the mind is the last refuge of mystery against the encroaching spread of science, and they don't like the idea of science engulfing the last bit of *terra incognita*. This means that they are threatened, I think irrationally, by the prospect that researchers in Artificial Intelligence may come to understand the human mind as well as biologists understand the genetic code, or as well as physicists understand electricity and magnetism. This could lead to the "evil scientist" (to take a stock character from science fiction) who can control you because he or she has a deep understanding of what's going on in your mind. This seems to me to be a totally valueless fear, one that you can set aside, for the simple reason that the human mind is full of an extraordinary amount of detailed knowledge, as, for example, Roger Schank has been pointing out.

As long as the scientist who is attempting to manipulate you does not share all your knowledge, his or her chances of manipulating you are minimal. People can always hit you over the head. They can do that now. We don't need Artificial Intelligence to manipulate people by putting them in chains or torturing them. But if someone tries to manipulate you by controlling your thoughts and ideas, that person will have to know what you know and more. The best way to keep yourself safe from that kind of manipulation is to be well informed.

Q: *Do you think we will be able to program self-consciousness into a computer?*

A: Yes, I do think that it's possible to program self-consciousness into a computer. *Self-consciousness* can mean many things. If you take the simplest, crudest notion of self-consciousness, I suppose that would be the sort of self-consciousness that a lobster has: When it's hungry, it eats something, but it never eats itself. It has some way of distinguishing between itself and the rest of the world, and it has a rather special regard for itself.

The lowly lobster is, in one regard, self-conscious. If you want to know whether or not you can create that on the computer, the answer is yes. It's no trouble at all. The computer is already a self-watching, self-monitoring sort of thing. That is an established part of the technology.

But, of course, most people have something more in mind when they speak of self-consciousness. It is that special inner light, that private way that it is with you that nobody else can share, something that is forever outside the bounds of computer science. How could a computer ever be conscious in this sense?

That belief, that very gripping, powerful intuition is, I think, in the end simply an illusion of common sense. It is as gripping as the common-sense illusion that the earth stands still and the sun goes around the earth. But the only way that those of us who do not believe in the illusion will ever convince the general public that it *is* an illusion is by gradually unfolding a very difficult and fascinating story about just what is going on in our minds.

In the interim, people like me—philosophers who have to live by our wits and tell a lot of stories—use what I call intuition pumps, little examples that help free up the imagination. I simply want to draw your attention to one fact. If you look at a computer—I don't care whether it's a giant Cray or a personal computer—if you open up the box and look inside and see those chips, you say, "No way could that be conscious. No way could that be self-conscious." But the same thing is true if you take the top off somebody's skull and look at the gray matter pulsing away in there. You think, "That is conscious? No way could that lump of stuff be conscious."

Of course, it makes no difference whether you look at it with a microscope or with a macroscope: At no level of inspection does a brain look like the seat of consciousness. Therefore, don't expect a computer to look like the seat of consciousness. If you want to get a grasp of how a computer could be conscious, it's no more difficult in the end than getting a grasp of how a brain could be conscious.

As we develop good accounts of consciousness, it will no longer seem so obvious to everyone that the idea of a self-conscious computer is a contradiction in terms. At the same time, I doubt that there will ever be self-conscious robots. But for boring reasons. There won't be any point in making them. Theoretically, could we make a gall bladder out of atoms? In principle we could. A gall bladder is just a collection of atoms, but manufacturing one would cost the moon. It would be more expensive than every project NASA has ever dreamed of, and there would be no scientific payoff. We wouldn't learn anything new about how gall bladders work. For the same reason, I don't think we're going to see really humanoid robots, because practical, cost-effective robots don't need to be very humanoid at all. They need to be like the robots you can already see at General Motors, or like boxy little computers that do special-purpose things.

The theoretical issues will be studied by artificial intelligence researchers by looking at models that, to the layman, will show very little sign of humanity at all, and it will be only by rather indirect arguments that anyone will be able to

appreciate that these models cast light on the deep theoretical question of how the mind is organized.

Postscript [1997]

In 1991, the First Annual Loebner Prize Competition was held in Boston at the Computer Museum. Hugh Loebner, a New York manufacturer, had put up the money for a prize—a bronze medal and $100,000—for the first computer program to pass the Turing test fair and square. The Prize Committee, of which I was Chairman until my resignation after the third competition, recognized that no program on the horizon could come close to passing the unrestricted test—the only test that is of any theoretical interest at all, as this essay has explained. So to make the competition interesting during the early years, some restrictions were adopted (and the award for winning the restricted test was dropped to $2,000). The first year there were ten terminals, with ten judges shuffling from terminal to terminal, each spending fifteen minutes in conversation with each terminal. Six of the ten contestants were programs, four were human "confederates" behind the scenes.

Each judge had to rank order all ten terminals from most human to least human. The winner of the restricted test would be the computer with the highest mean rating. The winning program would not have to fool any of the judges, nor would fooling a judge be in itself grounds for winning; highest mean ranking was all. But just in case some program *did* fool a judge, we thought this fact should be revealed, so judges were required to draw a line somewhere across their rank ordering, separating the humans from the machines.

We on the Prize Committee knew the low quality of the contesting programs that first year, and it seemed obvious to us that no program would be so lucky as to fool a single judge, but on the day of the competition, I got nervous. Just to be safe, I thought, we should have some certificate prepared to award to any programmer who happened to pull off this unlikely feat. While the press and the audience were assembling for the beginning of the competition, I rushed into a back room at the Computer Museum with a member of the staff and we cobbled up a handsome certificate with the aid of a handy desktop publisher. In the event, we had to hand out three of these certificates, for a total of seven positive misjudgments out of a possible sixty! The gullibility of the judges was simply astonishing to me. How *could* they have misjudged so badly? Here I had committed the sin I'd so often found in others: treating a failure of imagination as an insight into necessity. But remember that in order to make the competition much easier, we had tied the judges' hands in various ways—too many ways. The judges had been forbidden to *probe* the contestants aggressively, to conduct conversational experiments. (I may have chaired the committee, but I didn't always succeed in persuading a majority to adopt the rules I favored.) When the judges sat back passively, as instructed, and let the contestants lead them, they were readily taken in by the Potemkin village effect described in the essay.

None of the misjudgments counted as a real case of a computer passing the unrestricted Turing test, but they were still surprising to me. In the second year of the competition, we uncovered another unanticipated loophole: due to

faulty briefing of the confederates, several of them gave deliberately clunky, automaton-like answers. It turned out that they had decided to give the silicon contestants a sporting chance by acting as if they were programs! But once we'd straightened out these glitches in the rules and procedures, the competition worked out just as I had originally predicted: the computers stood out like sore thumbs even though there were still huge restrictions on topic. In the third year, two of the judges—journalists—each made a false *negative* judgment, declaring one of the less eloquent human confederates to be a computer. On debriefing, their explanation showed just how vast the gulf was between the computer programs and the people: they reasoned that the competition would not have been held if there weren't at least one halfway decent computer contestant, so they simply picked the least impressive human being and declared it to be a computer. But they could see the gap between the computers and the people as well as everybody else could.

The Loebner Prize Competition was a fascinating social experiment, and some day I hope to write up the inside story—a tale of sometimes hilarious misadventure, bizarre characters, interesting technical challenges, and more. But it never succeeded in attracting serious contestants from the world's best AI labs. Why not? In part because, as the essay argues, passing the Turing test is not a sensible research and development goal for serious AI. It requires too much Disney and not enough science. We might have corrected that flaw by introducing into the Loebner Competition something analogous to the "school figures" in ice-skating competition: theoretically interesting (but not crowd-pleasing) technical challenges such as parsing pronouns, or dealing creatively with enthymemes (arguments with unstated premises). Only those programs that performed well in the school figures—the serious competition—would be permitted into the final show-off round, where they could dazzle and amuse the onlookers with some cute Disney touches. Some such change in the rules would have wiped out all but the most serious and dedicated of the home hobbyists, and made the Loebner Competition worth winning (and not too embarrassing to lose). When my proposals along these lines were rejected, however, I resigned from the committee. The annual competitions continue, apparently, under the direction of Hugh Loebner. On the World Wide Web I just found the transcript of the conversation of the winning program in the 1996 completion. It was a scant improvement over 1991, still a bag of cheap tricks with no serious analysis of the meaning of the sentences. The Turing test is too difficult for the real world.

Notes

Originally appeared in Shafto, M., ed., *How We Know* (San Francisco: Harper & Row, 1985).

1. I thank Kenneth Colby for providing me with the complete transcripts (including the Judges' commentaries and reactions), from which these exchanges are quoted. The first published account of the experiment is Heiser et al. (1980, pp. 149–162). Colby (1981, pp. 515–560) discusses PARRY and its implications.

References

Block, N. (1982). "Psychologism and Behaviorism," *Philosophical Review*, 90, pp. 5–43.
Colby, K. M. (1981). "Modeling a Paranoid Mind," *Behavioral & Brain Sciences* 4 (4).

Descartes, R. (1637). *Discourse on Method*, LaFleur, Lawrence, trans., New York: Bobbs Merrill, 1960.

Haugeland, J. (1981). *Mind Design: Philosophy, Psychology, Artificial Intelligence,* Cambridge, MA: Bradford Books/MIT Press.

Heiser, J. F., Colby, K. M., Faught, W. S., and Parkinson, R. C. (1980). "Can Psychiatrists Distinguish Computer Simulation of Paranoia from the Real Thing? The Limitations of Turing-Like Tests as Measures of the Adequacy of Simulations," *Journal of Psychiatric Research* 15 (3).

Kolodner, J. L. (1983a). "Retrieval and Organization Strategies in a Conceptual Memory: A Computer Model" (Ph.D. diss.) Research Report #187, Dept. of Computer Science, Yale University.

Kolodner, J. L. (1983b). "Maintaining Organization in a Dynamic Long-term Memory," *Cognitive Science* 7.

Kolodner, J. L. (1983c). "Reconstructive Memory: A Computer Model," *Cognitive Science* 7.

Turing, A. (1950). Computing machinery and intelligence. *Mind,* 59 (236), pp. 433–460. Reprinted in Hofstadter, D., and Dennett, D. C., eds., *The Mind's I* (New York: Basic Books, 1981), pp. 54–67.

Weizenbaum, J. (1974). Letter to the editor. *Communications of the Association for Computing Machinery,* 17 (9) (September).

Winograd, T. (1972). *Understanding Natural Language*, New York: Academic Press.

PART II

Neural Networks

Chapter 4

The Appeal of Parallel Distributed Processing

Jay L. McClelland, David E. Rumelhart, and Geoffrey E. Hinton

---

What makes people smarter than machines? They certainly are not quicker or more precise. Yet people are far better at perceiving objects in natural scenes and noting their relations, at understanding language and retrieving contextually appropriate information from memory, at making plans and carrying out contextually appropriate actions, and at a wide range of other natural cognitive tasks. People are also far better at learning to do these things more accurately and fluently through processing experience.

What is the basis for these differences? One answer, perhaps the classic one we might expect from artificial intelligence, is "software." If we only had the right computer program, the argument goes, we might be able to capture the fluidity and adaptability of human information processing.

Certainly this answer is partially correct. There have been great breakthroughs in our understanding of cognition as a result of the development of expressive high-level computer languages and powerful algorithms. No doubt there will be more such breakthroughs in the future. However, we do not think that software is the whole story.

In our view, people are smarter than today's computers because the brain employs a basic computational architecture that is more suited to deal with a central aspect of the natural information processing tasks that people are so good at. In this chapter, we will show through examples that these tasks generally require the simultaneous consideration of many pieces of information or constraints. Each constraint may be imperfectly specified and ambiguous, yet each can play a potentially decisive role in determining the outcome of processing. After examining these points, we will introduce a computational framework for modeling cognitive processes that seems well suited to exploiting these constaints and that seems closer than other frameworks to the style of computation as it might be done by the brain. We will review several early examples of models developed in this framework, and we will show that the mechanisms these models employ can give rise to powerful emergent properties that begin to suggest attractive alternatives to traditional accounts of various aspects of cognition. We will also show that models of this class provide a basis for understanding how learning can occur spontaneously, as a by-product of processing activity.

From chapter 1 in *Parallel Distributed Processing*, Vol. 1: *Foundations*, ed. D. E. Rumelhart, J. L. McClelland, and the PDP Research Group (Cambridge, MA: MIT Press, 1986), 3–44. Reprinted with permission.

Multiple Simultaneous Constraints

Reaching and Grasping Hundreds of times each day we reach for things. We nearly never think about these acts of reaching. And yet, each time, a large number of different considerations appear to jointly determine exactly how we will reach for the object. The position of the object, our posture at the time, what else we may also be holding, the size, shape, and anticipated weight of the object, any obstacles that may be in the way—all of these factors jointly determine the exact method we will use for reaching and grasping.

Consider the situation shown in figure 4.1. Figure 4.1A shows Jay McClelland's hand, in typing position at his terminal. Figure 4.1B indicates the position his hand assumed in reaching for a small knob on the desk beside the terminal. We will let him describe what happened in the first person:

> On the desk next to my terminal are several objects—a chipped coffee mug, the end of a computer cable, a knob from a clock radio. I decide to pick the knob up. At first I hesitate, because it doesn't seem possible. Then I just reach for it, and find myself grasping the knob in what would normally be considered a very awkward position—but it solves all of the constraints. I'm not sure what all the details of the movement were, so I let myself try it a few times more. I observe that my right hand is carried up off the keyboard, bent at the elbow, until my forearm is at about a 30° angle to the desk top and parallel to the side of the terminal. The palm is facing downward through most of this. Then, my arm extends and lowers down more or less parallel to the edge of the desk and parallel to the side of the terminal and, as it drops, it turns about 90° so that the palm is facing the cup and the thumb and index finger are below. The turning motion occurs just in time, as my hand drops, to avoid hitting the coffee cup. My index finger and thumb close in on the knob and grasp it, with my hand completely upside down.

Though the details of what happened here might be quibbled with, the broad outlines are apparent. The shape of the knob and its position on the table; the starting position of the hand on the keyboard; the positions of the terminal, the cup, and the knob; and the constraints imposed by the structure of the arm and the musculature used to control it—all these things conspired to lead to a solution which exactly suits the problem. If any of these constraints had not been included, the movement would have failed. The hand would have hit the cup or the terminal—or it would have missed the knob.

The Mutual Influence of Syntax and Semantics Multiple constraints operate just as strongly in language processing as they do in reaching and grasping. Rumelhart (1977) has documented many of these multiple constraints. Rather than catalog them here, we will use a few examples from language to illustrate the fact that the constraints tend to be reciprocal: The example shows that they do not run only from syntax to semantics—they also run the other way.

It is clear, of course, that syntax constrains the assignment of meaning. Without the syntactic rules of English to guide us, we cannot correctly understand who has done what to whom in the following sentence:

A

B

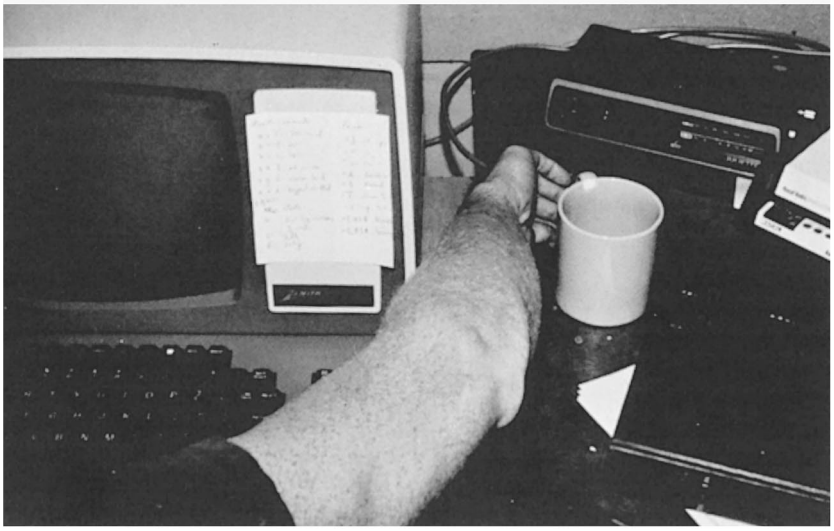


Figure 4.1
A: An everyday situation in which it is necessary to take into account a large number of constraints to grasp a desired object. In this case the target object is the small knob to the left of the cup. B: The posture the arm arrives at in meeting these constraints.

The boy the man chased kissed the girl.

But consider these examples (Rumelhart, 1977; Schank, 1973):

I saw the Grand Canyon flying to New York.
I saw the sheep grazing in the field.

Our knowledge of syntactic rules alone does not tell us what grammatical role is played by the prepositional phrases in these two cases. In the first, "flying to New York" is taken as describing the context in which the speaker saw the Grand Canyon—while he was flying to New York. In the second, "grazing in the field" could syntactically describe an analogous situation, in which the speaker is grazing in the field, but this possibility does not typically become available on first reading. Instead we assign "grazing in the field" as a modifier of the sheep (roughly, "who were grazing in the field"). The syntactic structure of each of these sentences, then, is determined in part by the semantic relations that the constituents of the sentence might plausibly bear to one another. Thus, the influences appear to run both ways, from the syntax to the semantics and from the semantics to the syntax.

In these examples, we see how syntactic considerations influence semantic ones and how semantic ones influence syntactic ones. We cannot say that one kind of constraint is primary.

Mutual constraints operate, not only between syntactic and semantic processing, but also within each of these domains as well. Here we consider an example from syntactic processing, namely, the assignment of words to syntactic categories. Consider the sentences:

I like the joke.
I like the drive.
I like to joke.
I like to drive.

In this case it looks as though the words *the* and *to* serve to determine whether the following word will be read as a noun or a verb. This, of course, is a very strong constraint in English and can serve to force a verb interpretation of a word that is not ordinarily used this way:

I like to mud.

On the other hand, if the information specifying whether the function word preceding the final word is *to* or *the* is ambiguous, then the typical reading of the word that follows it will determine which way the function word is heard. This was shown in an experiment by Isenberg, Walker, Ryder, and Schweikert (1980). They presented sounds halfway between *to* (actually/tˆ/) and *the* (actually/ðˆ/) and found that words like *joke*, which we tend to think of first as nouns, made subjects hear the marginal stimuli as *the*, while words like *drive*, which we tend to think of first as verbs, made subjects hear the marginal stimuli as *to*. Generally, then, it would appear that each word can help constrain the syntactic role, and even the identity, of every other word.

Simultaneous Mutual Constraints in Word Recognition Just as the syntactic role of one word can influence the role assigned to another in analyzing sentences,

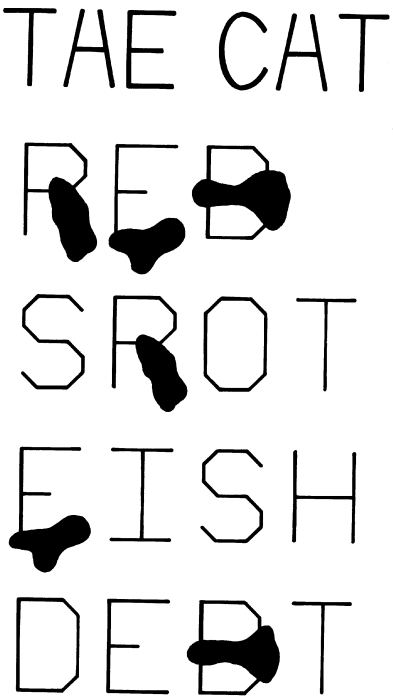


Figure 4.2
Some ambiguous displays. The first one is from Selfridge, 1955. The second line shows that three ambiguous characters can each constrain the identity of the others. The third, fourth, and fifth lines show that these characters are indeed ambiguous in that they assume other identities in other contexts. (The ink-blot technique of making letters ambiguous is due to Lindsay and Norman, 1972).

so the identity of one letter can influence the identity assigned to another in reading. A famous example of this, from Selfridge, is shown in figure 4.2. Along with this is a second example in which none of the letters, considered separately, can be identified unambiguously, but in which the possibilities that the visual information leaves open for each so constrain the possible identities of the others that we are capable of identifying all of them.

At first glance, the situation here must seem paradoxical: The identity of each letter is constrained by the identities of each of the others. But since in general we cannot know the identities of any of the letters until we have established the identities of the others, how can we get the process started?

The resolution of the paradox, of course, is simple. One of the different possible letters in each position fits together with the others. It appears then that our perceptual system is capable of exploring all these possibilities without committing itself to one until all of the constraints are taken into account.

Understanding through the Interplay of Multiple Sources of Knowledge It is clear that we know a good deal about a large number of different standard situations. Several theorists have suggested that we store this knowledge in terms of structures called variously: *scripts* (Schank, 1976), *frames* (Minsky, 1975), or

schemata (Norman & Bobrow, 1976; Rumelhart, 1975). Such knowledge structures are assumed to be the basis of comprehension. A great deal of progress has been made within the context of this view.

However, it is important to bear in mind that most everyday situations cannot be rigidly assigned to just a single script. They generally involve an interplay between a number of different sources of information. Consider, for example, a child's birthday party at a restaurant. We know things about birthday parties, and we know things about restaurants, but we would not want to assume that we have explicit knowledge (at least, not in advance of our first restaurant birthday party) about the conjunction of the two. Yet we can imagine what such a party might be like. The fact that the party was being held in a restaurant would modify certain aspects of our expectations for birthday parties (we would not expect a game of Pin-the-Tail-on-the-Donkey, for example), while the fact that the event was a birthday party would inform our expectations for what would be ordered and who would pay the bill.

Representations like scripts, frames, and schemata are useful structures for encoding knowledge, although we believe they only approximate the underlying structure of knowledge representation that emerges from the class of models we consider in this chapter. Our main point here is that any theory that tries to account for human knowledge using script-like knowledge structures will have to allow them to interact with each other to capture the generative capacity of human understanding in novel situations. Achieving such interactions has been one of the greatest difficulties associated with implementing models that really think generatively using script- or frame-like representations.

Parallel Distributed Processing

In the examples we have considered, a number of different pieces of information must be kept in mind at once. Each plays a part, constraining others and being constrained by them. What kinds of mechanisms seem well suited to these task demands? Intuitively, these tasks seem to require mechanisms in which each aspect of the information in the situation can act on other aspects, simultaneously influencing other aspects and being influenced by them. To articulate these intuitions, we and others have turned to a class of models we call *Parallel Distributed Processing* (PDP) models. These models assume that information processing takes place through the interactions of a large number of simple processing elements called units, each sending excitatory and inhibitory signals to other units. In some cases, the units stand for possible hypotheses about such things as the letters in a particular display or the syntactic roles of the words in a particular sentence. In these cases, the activations stand roughly for the strengths associated with the different possible hypotheses, and the interconnections among the units stand for the constraints the system knows to exist between the hypotheses. In other cases, the units stand for possible goals and actions, such as the goal of typing a particular letter, or the action of moving the left index finger, and the connections relate goals to subgoals, subgoals to actions, and actions to muscle movements. In still other cases, units stand not for particular hypotheses or goals, but for aspects of these things. Thus a

Figure 4.3
The arborizations of about 1 percent of the neurons near a vertical slice through the cerebral cortex. The full height of the figure corresponds to the thickness of the cortex, which is in this instance about 2 mm. (From *Mechanics of the Mind*, p. 84, by C. Blakemore, 1977, Cambridge, England: Cambridge University Press. Copyright 1977 by Cambridge University Press. Reprinted with permission.)

hypothesis about the identity of a word, for example, is itself distributed in the activations of a large number of units.

PDP Models: Cognitive Science or Neuroscience?

One reason for the appeal of PDP models is their obvious "physiological" flavor: They seem so much more closely tied to the physiology of the brain than are other kinds of information-processing models. The brain consists of a large number of highly interconnected elements (figure 4.3) which apparently send very simple excitatory and inhibitory messages to each other and update their excitations on the basis of these simple messages. The properties of the units in many PDP models were inspired by basic properties of the neural hardware.

Though the appeal of PDP models is definitely enhanced by their physiological plausibility and neural inspiration, these are not the primary bases for their appeal to us. We are, after all, cognitive scientists, and PDP models appeal to us for psychological and computational reasons. They hold out the hope of offering computationally sufficient and psychologically accurate mechanistic accounts of the phenomena of human cognition which have eluded successful explication in conventional computational formalisms; and they have radically

altered the way we think about the time-course of processing, the nature of representation, and the mechanisms of learning.

The Microstructure of Cognition

The process of human cognition, examined on a time scale of seconds and minutes, has a distinctly sequential character to it. Ideas come, seem promising, and then are rejected; leads in the solution to a problem are taken up, then abandoned and replaced with new ideas. Though the process may not be discrete, it has a decidedly sequential character, with transitions from state-to-state occurring, say, two or three times a second. Clearly, any useful description of the overall organization of this sequential flow of thought will necessarily describe a sequence of states.

But what is the internal structure of each of the states in the sequence, and how do they come about? Serious attempts to model even the simplest macrosteps of cognition—say, recognition of single words—require vast numbers of microsteps if they are implemented sequentially. As Feldman and Ballard (1982) have pointed out, the biological hardware is just too sluggish for sequential models of the microstructure to provide a plausible account, at least of the microstructure of *human* thought. And the time limitation only gets worse, not better, when sequential mechanisms try to take large numbers of constraints into account. Each additional constraint requires more time in a sequential machine, and, if the constraints are imprecise, the constraints can lead to a computational explosion. Yet people get faster, not slower, when they are able to exploit additional constraints.

Parallel distributed processing models offer alternatives to serial models of the microstructure of cognition. They do not deny that there is a macrostructure, just as the study of subatomic particles does not deny the existence of interactions between atoms. What PDP models do is describe the internal structure of the larger units, just as subatomic physics describes the internal structure of the atoms that form the constituents of larger units of chemical structure.

The analysis of the microstructure of cognition has important implications for most of the central issues in cognitive science. In general, from the PDP point of view, the objects referred to in macrostructural models of cognitive processing are seen as approximate descriptions of emergent properties of the microstructure. Sometimes these approximate descriptions may be sufficiently accurate to capture a process or mechanism well enough; but many times, we will argue, they fail to provide sufficiently elegant or tractable accounts that capture the very flexibility and open-endedness of cognition that their inventors had originally intended to capture. We hope that our analysis of PDP models will show how an examination of the microstructure of cognition can lead us closer to an adequate description of the real extent of human processing and learning capacities.

The development of PDP models is still in its infancy. Thus far the models which have been proposed capture simplified versions of the kinds of phenomena we have been describing rather than the full elaboration that these phenomena display in real settings. But we think there have been enough steps forward to warrant a concerted effort at describing where the approach has

gotten and where it is going now, and to point out some directions for the future.

The rest of this chapter attempts to describe in informal terms a number of the models which have been proposed in previous work and to show that the approach is indeed a fruitful one. It also contains a brief description of the major sources of the inspiration we have obtained from the work of other researchers.

Examples of PDP Models

In what follows, we review a number of recent applications of PDP models to problems in motor control, perception, memory, and language. In many cases, as we shall see, parallel distributed processing mechanisms are used to provide natural accounts of the exploitation of multiple, simultaneous, and often mutual constraints. We will also see that these same mechanisms exhibit emergent properties which lead to novel interpretations of phenomena which have traditionally been interpreted in other ways.

Motor Control

Having started with an example of how multiple constraints appear to operate in motor programming, it seems appropriate to mention two models in this domain. These models have not developed far enough to capture the full details of obstacle avoidance and multiple constraints on reaching and grasping, but there have been applications to two problems with some of these characteristics.

Finger Movements in Skilled Typing One might imagine, at first glance, that typists carry out keystrokes successively, first programming one stroke and then, when it is completed, programming the next. However, this is not the case. For skilled typists, the fingers are continually anticipating upcoming keystrokes. Consider the word *vacuum*. In this word, the *v*, *a*, and *c* are all typed with the left hand, leaving the right hand nothing to do until it is time to type the first *u*. However, a high speed film of a good typist shows that the right hand moves up to anticipate the typing of the *u*, even as the left hand is just beginning to type the *v*. By the time the *c* is typed the right index finger is in position over the *u* and ready to strike it.

When two successive key strokes are to be typed with the fingers of the same hand, concurrent preparation to type both can result in similar or conflicting instructions to the fingers and/or the hand. Consider, in this light, the difference between the sequence *ev* and the sequence *er*. The first sequence requires the typist to move up from home row to type the *e* and to move down from the home row to type the *v*, while in the second sequence, both the *e* and the *r* are above the home row.

The hands take very different positions in these two cases. In the first case, the hand as a whole stays fairly stationary over the home row. The middle finger moves up to type the *e*, and the index finger moves down to type the *v*. In the second case, the hand as a whole moves up, bringing the middle finger over the *e* and the index finger over the *r*. Thus, we can see that several letters can simultaneously influence the positioning of the fingers and the hands.

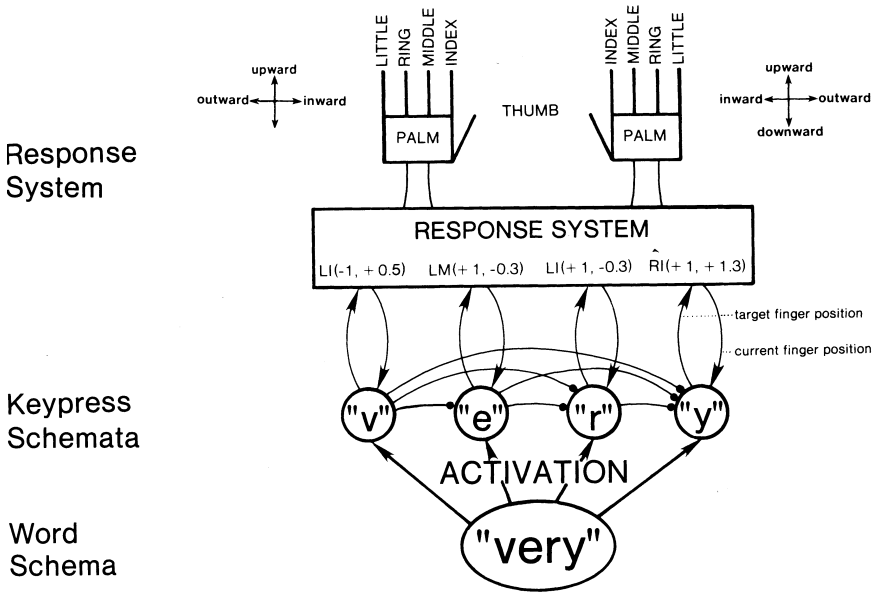


Figure 4.4
The interaction of activations in typing the word *very*. The *very* unit is activated from outside the model. It in turn activates the units for each of the component letters. Each letter unit specifies the target finger positions, specified in a keyboard coordinate system. L and R stand for the left and right hands, and I and M for the index and middle fingers. The letter units receive information about the current finger position from the response system. Each letter unit inhibits the activation of all letter units that follow it in the word: inhibitory connections are indicated by the lines with solid dots at their terminations. (From "Simulating a Skilled Typist: A Study of Skilled Motor Performance" by D. E. Rumelhart and D. A. Norman, 1982, *Cognitive Science, 6*, p. 12. Copyright 1982 by Ablex Publishing. Reprinted with permission.)

From the point of view of optimizing the efficiency of the typing motion, these different patterns seem very sensible. In the first case, the hand as a whole is maintained in a good compromise position to allow the typist to strike both letters reasonably efficiently by extending the fingers up or down. In the second case, the need to extend the fingers is reduced by moving the whole hand up, putting it in a near-optimal position to strike either key.

Rumelhart and Norman (1982) have simulated these effects using PDP mechanisms. Figure 4.4 illustrates aspects of the model as they are illustrated in typing the word *very*. In brief, Rumelhart and Norman assumed that the decision to type a word caused activation of a unit for that word. That unit, in turn, activated units corresponding to each of the letters in the word. The unit for the first letter to be typed was made to inhibit the units for the second and following letters, the unit for the second to inhibit the third and following letters, and so on. As a result of the interplay of activation and inhibition among these units, the unit for the first letter was at first the most strongly active, and the units for the other letters were partially activated.

Each letter unit exerts influences on the hand and finger involved in typing the letter. The *v* unit, for example, tends to cause the index finger to move down and to cause the whole hand to move down with it. The *e* unit, on the

other hand, tends to cause the middle finger on the left hand to move up and to cause the whole hand to move up also. The *r* unit also causes the left index finger to move up and the left hand to move up with it.

The extent of the influences of each letter on the hand and finger it directs depends on the extent of the activation of the letter. Therefore, at first, in typing the word *very*, the *v* exerts the greatest control. Because the *e* and *r* are simultaneously pulling the hand up, though, the *v* is typed primarily by moving the index finger, and there is little movement on the whole hand.

Once a finger is within a certain striking distance of the key to be typed, the actual pressing movement is triggered, and the keypress occurs. The keypress itself causes a strong inhibitory signal to be sent to the unit for the letter just typed, thereby removing this unit from the picture and allowing the unit for the next letter in the word to become the most strongly activated.

This mechanism provides a simple way for all of the letters to jointly determine the successive configurations the hand will enter into in the process of typing a word. This model has shown considerable success predicting the time between successive keystrokes as a function of the different keys involved. Given a little noise in the activation process, it can also account for some of the different kinds of errors that have been observed in transcription typing.

The typing model represents an illustration of the fact that serial behavior—a succession of key strokes—is not necessarily the result of an inherently serial processing mechanism. In this model, the sequential structure of typing emerges from the interaction of the excitatory and inhibitory influences among the processing units.

Reaching for an Object without Falling Over Similar mechanisms can be used to model the process of reaching for an object without losing one's balance while standing, as Hinton (1984) has shown. He considered a simple version of this task using a two-dimensional "person" with a foot, a lower leg, an upper leg, a trunk, an upper arm, and a lower arm. Each of these limbs is joined to the next at a joint which has a single degree of rotational freedom. The task posed to this person is to reach a target placed somewhere in front of it, without taking any steps and without falling down. This is a simplified version of the situation in which a real person has to reach out in front for an object placed somewhere in the plane that vertically bisects the body. The task is not as simple as it looks, since if we just swing an arm out in front of ourselves, it may shift our center of gravity so far forward that we will lose our balance. The problem, then, is to find a set of joint angles that simultaneously solves the two constraints on the task. First, the tip of the forearm must touch the object. Second, to keep from falling down, the person must keep its center of gravity over the foot.

To do this, Hinton assigned a single processor to each joint. On each computational cycle, each processor received information about how far the tip of the hand was from the target and where the center of gravity was with respect to the foot. Using these two pieces of information, each joint adjusted its angle so as to approach the goals of maintaining balance and bringing the tip closer to the target. After a number of iterations, the stick-person settled on postures that satisfied the goal of reaching the target and the goal of maintaining the center of gravity over the "feet."

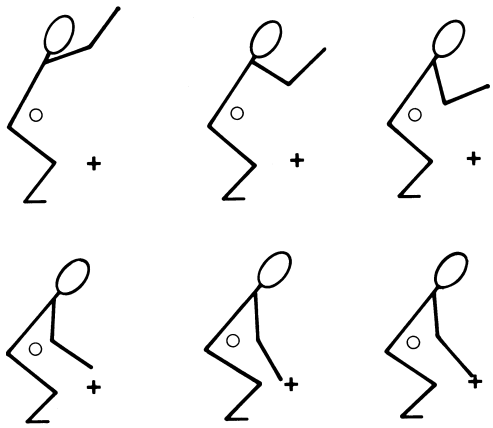


Figure 4.5
A sequence of configurations assumed by the stick "person" performing the reaching task described in the text, from Hinton (1984). The small circle represents the center of gravity of the whole stick-figure, and the cross represents the goal to be reached. The configuration is shown on every second iteration.

Though the simulation was able to perform the task, eventually satisfying both goals at once, it had a number of inadequacies stemming from the fact that each joint processor attempted to achieve a solution in ignorance of what the other joints were attempting to do. This problem was overcome by using additional processors responsible for setting combinations of joint angles. Thus, a processor for flexion and extension of the leg would adjust the knee, hip, and ankle joints synergistically, while a processor for flexion and extension of the arm would adjust the shoulder and elbow together. With the addition of processors of this form, the number of iterations required to reach a solution was greatly reduced, and the form of the approach to the solution looked very natural. The sequence of configurations attained in one processing run is shown in figure 4.5.

Explicit attempts to program a robot to cope with the problem of maintaining balance as it reaches for a desired target have revealed the difficulty of deriving explicitly the right combinations of actions for each possible starting state and goal state. This simple model illustrates that we may be wrong to seek such an explicit solution. We see here that a solution to the problem can emerge from the action of a number of simple processors each attempting to honor the constraints independently.

Perception

Stereoscopic Vision One early model using parallel distributed processing was the model of stereoscopic depth perception proposed by Marr and Poggio (1976). Their theory proposed to explain the perception of depth in random-dot stereograms (Julesz, 1971; see figure 4.6) in terms of a simple distributed processing mechanism.

Julesz's random-dot stereograms present interesting challenges to mechanisms of depth perception. A stereogram consists of two random-dot patterns.

Figure 4.6
Random-dot stereograms. The two patterns are identical except that the pattern of dots in the central region of the left pattern are shifted over with respect to those in the right. When viewed stereoscopically such that the left pattern projects to the left eye and the right pattern to the right eye, the shifted area appears to hover above the page. Some readers may be able to achieve this by converging to a distant point (e.g., a far wall) and then interposing the figure into the line of sight. (From *Foundations of Cyclopean Perception*, p. 21, by B. Julesz, 1971, Chicago: University of Chicago Press. Copyright 1971 by Bell Telephone Laboratories, Inc. Reprinted by permission.)

In a simple stereogram such as the one shown here, one pattern is an exact copy of the other except that the pattern of dots in a region of one of the patterns is shifted horizontally with respect to the rest of the pattern. Each of the two patterns—corresponding to two retinal images—consists entirely of a pattern of random dots, so there is no information in either of the two views considered alone that can indicate the presence of different surfaces, let alone depth relations among those surfaces. Yet, when one of these dot patterns is projected to the left eye and the other to the right eye, an observer sees each region as a surface, with the shifted region hovering in front of or behind the other, depending on the direction of the shift.

What kind of a mechanism might we propose to account for these facts? Marr and Poggio (1976) began by explicitly representing the two views in two arrays, as human observers might in two different retinal images. They noted that corresponding black dots at different perceived distances from the observer will be offset from each other by different amounts in the two views. The job of the model is to determine which points correspond. This task is, of course, made difficult by the fact that there will be a very large number of spurious correspondences of individual dots. The goal of the mechanism, then, is to find those correspondences that represent real correspondences in depth and suppress those that represent spurious correspondences.

To carry out this task, Marr and Poggio assigned a processing unit to each possible conjunction of a point in one image and a point in the other. Since the eyes are offset horizontally, the possible conjunctions occur at various offsets or disparities along the horizontal dimension. Thus, for each point in one eye, there was a set of processing units with one unit assigned to the conjunction of that point and the point at each horizontal offset from it in the other eye.

Each processing unit received activation whenever both of the points the unit stood for contained dots. So far, then, units for both real and spurious correspondences would be equally activated. To allow the mechanism to find the right correspondences, they pointed out two general principles about the visual world: (a) Each point in each view generally corresponds to one and only one point in the other view, and (b) neighboring points in space tend to be at nearly the same depth and therefore at about the same disparity in the two images. While there are discontinuities at the edges of things, over most of a two-dimensional view of the world there will be continuity. These principles are called the *uniqueness* and *continuity* constraints, respectively.

Marr and Poggio incorporated these principles into the interconnections between the processing units. The uniqueness constraint was captured by inhibitory connections among the units that stand for alternative correspondences of the same dot. The continuity principle was captured by excitatory connections among the units that stand for similar offsets of adjacent dots.

These additional connections allow the Marr and Poggio model to "solve" stereograms like the one shown in the figure. At first, when a pair of patterns is presented, the units for all possible correspondences of a dot in one eye with a dot in the other will be equally excited. However, the excitatory connections cause the units for the correct conjunctions to receive more excitation than units for spurious conjunctions, and the inhibitory connections allow the units for the correct conjunctions to turn off the units for the spurious connections. Thus, the model tends to settle down into a stable state in which only the correct correspondence of each dot remains active.

There are a number of reasons why Marr and Poggio (1979) modified this model (see Marr, 1982, for a discussion), but the basic mechanisms of mutual excitation between units that are mutually consistent and mutual inhibition between units that are mutually incompatible provide a natural mechanism for settling on the right conjunctions of points and rejecting spurious ones. The model also illustrates how general principles or rules such as the uniqueness and continuity principles may be embodied in the connections between processing units, and how behavior in accordance with these principles can emerge from the interactions determined by the pattern of these interconnections.

Perceptual Completion of Familiar Patterns Perception, of course, is influenced by familiarity. It is a well-known fact that we often misperceive unfamiliar objects as more familiar ones and that we can get by with less time or with lower-quality information in perceiving familiar items than we need for perceiving unfamiliar items. Not only does familiarity help us determine what the higher-level structures are when the lower-level information is ambiguous; it also allows us to fill in missing lower-level information within familiar higher-order patterns. The well-known *phonemic restoration effect* is a case in point. In this phenomenon, perceivers hear sounds that have been cut out of words as if they had actually been present. For example, Warren (1970) presented *legi-#lature* to subjects, with a click in the location marked by the #. Not only did subjects correctly identify the word legislature; they also heard the missing /s/ just as though it had been presented. They had great difficulty localizing the click, which they tended to hear as a disembodied sound. Similar phenomena

have been observed in visual perception of words since the work of Pillsbury (1897).

Two of us have proposed a model describing the role of familiarity in perception based on excitatory and inhibitory interactions among units standing for various hypotheses about the input at different levels of abstraction (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). The model has been applied in detail to the role of familiarity in the perception of letters in visually presented words, and has proved to provide a very close account of the results of a large number of experiments.

The model assumes that there are units that act as detectors for the visual features which distinguish letters, with one set of units assigned to detect the features in each of the different letter-positions in the word. For four-letter words, then, there are four such sets of detectors. There are also four sets of detectors for the letters themselves and a set of detectors for the words.

In the model, each unit has an activation value, corresponding roughly to the strength of the hypothesis that what that unit stands for is present in the perceptual input. The model honors the following important relations which hold between these "hypotheses" or activations: First, to the extent that two hypotheses are mutually consistent, they should support each other. Thus, units that are mutually consistent, in the way that the letter *T* in the first position is consistent with the word *TAKE*, tend to excite each other. Second, to the extent that two hypotheses are mutually inconsistent, they should weaken each other. Actually, we can distinguish two kinds of inconsistency: The first kind might be called between-level inconsistency. For example, the hypothesis that a word begins with a *T* is inconsistent with the hypothesis that the word is *MOVE*. The second might be called mutual exclusion. For example, the hypothesis that a word begins with *T* excludes the hypothesis that it begins with *R* since a word can only begin with one letter. Both kinds of inconsistencies operate in the word perception model to reduce the activations of units. Thus, the letter units in each position compete with all other letter units in the same position, and the word units compete with each other. This type of inhibitory interaction is often called *competitive inhibition*. In addition, there are inhibitory interactions between incompatible units on different levels. This type of inhibitory interaction is simply called *between-level inhibition*.

The set of excitatory and inhibitory interactions between units can be diagrammed by drawing excitatory and inhibitory links between them. The whole picture is too complex to draw, so we illustrate only with a fragment: Some of the interactions between some of the units in this model are illustrated in figure 4.7.

Let us consider what happens in a system like this when a familiar stimulus is presented under degraded conditions. For example, consider the display shown in figure 4.8. This display consists of the letters *W*, *O*, and *R*, completely visible, and enough of a fourth letter to rule out all letters other than *R* and *K*. Before onset of the display, the activations of the units are set at or below 0. When the display is presented, detectors for the features present in each position become active (i.e., their activations grow above 0). At this point, they begin to excite and inhibit the corresponding detectors for letters. In the first three positions, *W*, *O*, and *R* are unambiguously activated, so we will focus our

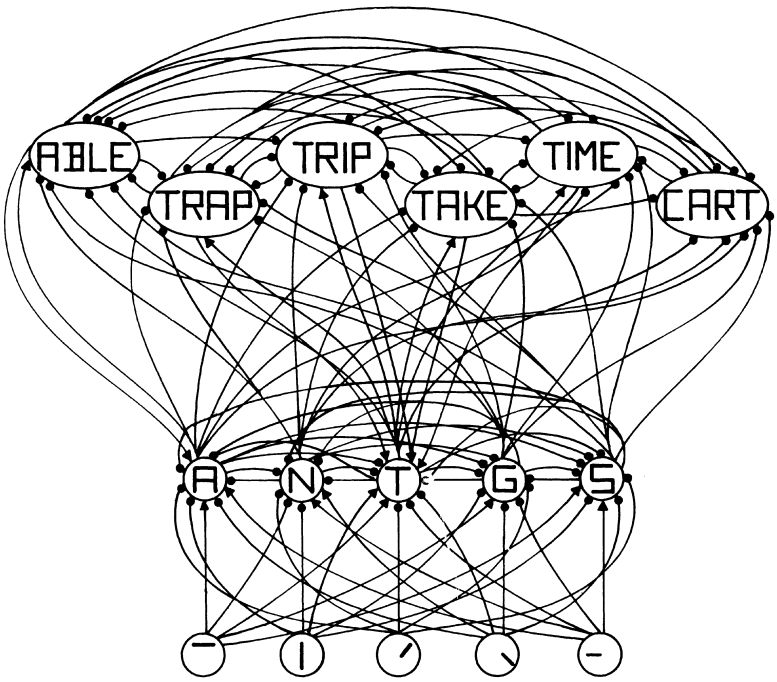


Figure 4.7
The unit for the letter *T* in the first position of a four-letter array and some of its neighbors. Note that the feature and letter units stand only for the first position; in a complete picture of the units needed from processing four-letter displays, there would be four full sets of feature detectors and four full sets of letter detectors. (From "An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings" by J. L. McClelland and D. E. Rumelhart, 1981, *Psychological Review, 88*, p. 380. Copyright 1981 by the American Psychological Association. Reprinted by permission.)

attention on the fourth position where *R* and *K* are both equally consistent with the active features. Here, the activations of the detectors for *R* and *K* start out growing together, as the feature detectors below them become activated. As these detectors become active, they and the active letter detectors for *W*, *O*, and *R* in the other positions start to activate detectors for words which have these letters in them and to inhibit detectors for words which do not have these letters. A number of words are partially consistent with the active letters, and receive some net excitation from the letter level, but only the word *WORK* matches one of the active letters in all four positions. As a result, *WORK* becomes more active than any other word and inhibits the other words, thereby successfully dominating the pattern of activation among the word units. As it grows in strength, it sends feedback to the letter level, reinforcing the activations of the *W*, *O*, *R*, and *K* in the corresponding positions. In the fourth position, this feedback gives *K* the upper hand over *R*, and eventually the stronger activation of the *K* detector allows it to dominate the pattern of activation, suppressing the *R* detector completely.

This example illustrates how PDP models can allow knowledge about what letters go together to form words to work together with natural constraints on

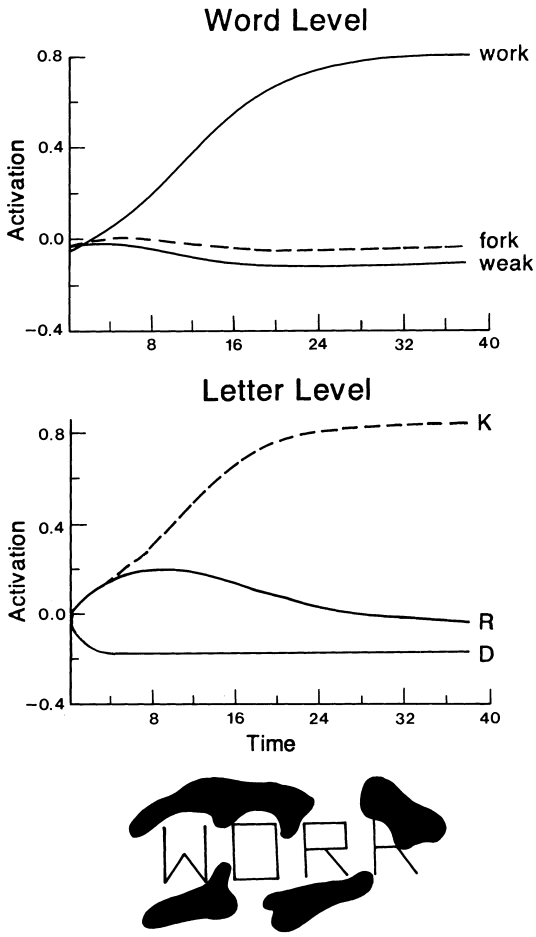


Figure 4.8
A possible display which might be presented to the interactive activation model of word recognition, and the resulting activations of selected letter and word units. The letter units are for the letters indicated in the fourth position of a four-letter display.

the task (i.e., that there should only be one letter in one place at one time), to produce perceptual completion in a simple and direct way.

Completion of Novel Patterns However, the perceptual intelligence of human perceivers far exceeds the ability to recognize familiar patterns and fill in missing portions. We also show facilitation in the perception of letters in unfamiliar letter strings which are word-like but not themselves actually familiar.

One way of accounting for such performances is to imagine that the perceiver possesses, in addition to detectors for familiar words, sets of detectors for regular subword units such as familiar letter clusters, or that they use abstract rules, specifying which classes of letters can go with which others in different contexts. It turns out, however, that the model we have already described needs no such additional structure to produce perceptual facilitation for word-like letter strings; to this extent it acts as if it "knows" the orthographic

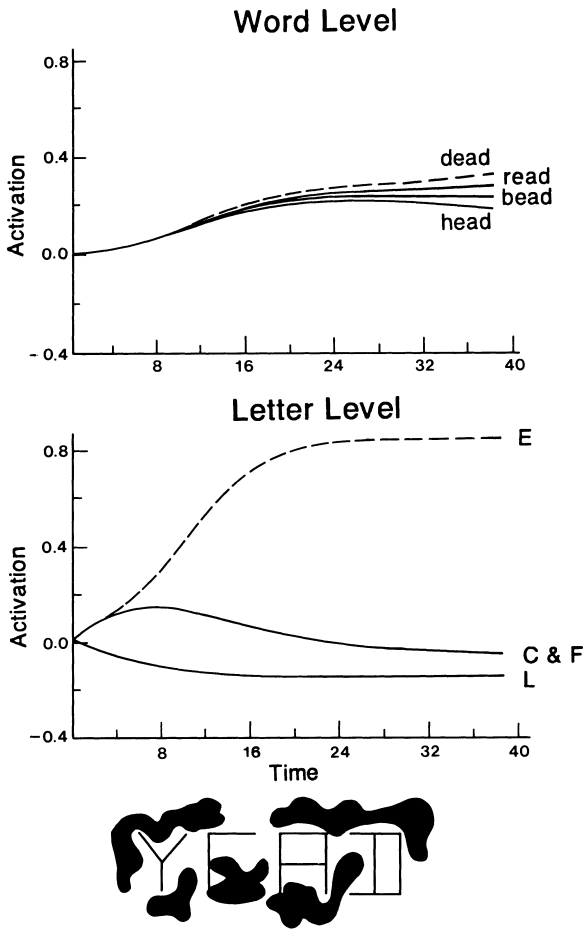


Figure 4.9
An example of a nonword display that might be presented to the interactive activation model of word recognition and the response of selected units at the letter and word levels. The letter units illustrated are detectors for letters in the second input position.

structure of English. We illustrate this feature of the model with the example shown in figure 4.9, where the nonword *YEAD* is shown in degraded form so that the second letter is incompletely visible. Given the information about this letter, considered alone, either *E* or *F* would be possible in the second position. Yet our model will tend to complete this letter as an *E*.

The reason for this behavior is that, when *YEAD* is shown, a number of words are partially activated. There is no word consistent with *Y*, *E* or *F*, *A*, and *D*, but there are words which match *YEA_* (*YEAR*, for example) and others which match *_EAD* (*BEAD, DEAD, HEAD*, and *READ*, for example). These and other near misses are partially activated as a result of the pattern of activation at the letter level. While they compete with each other, none of these words gets strongly enough activated to completely suppress all the others. Instead, these units act as a group to reinforce particularly the letters *E* and *A*. There are

no close partial matches which include the letter *F* in the second position, so this letter receives no feedback support. As a result, *E* comes to dominate, and eventually suppress, the *F* in the second position.

The fact that the word perception model exhibits perceptual facilitation to pronounceable nonwords as well as words illustrates once again how behavior in accordance with general principles or rules can emerge from the interactions of simple processing elements. Of course, the behavior of the word perception model does not implement exactly any of the systems of orthographic rules that have been proposed by linguists (Chomsky & Halle, 1968; Venesky, 1970) or psychologists (Spoehr & Smith, 1975). In this regard, it only approximates such rule-based descriptions of perceptual processing. However, rule systems such as Chomsky and Halle's or Venesky's appear to be only approximately honored in human performance as well (Smith & Baker, 1976). Indeed, some of the discrepancies between human performance data and rule systems occur in exactly the ways that we would predict from the word perception model (Rumelhart & McClelland, 1982). This illustrates the possibility that PDP models may provide more accurate accounts of the details of human performance than models based on a set of rules representing human competence—at least in some domains.

Retrieving Information from Memory

Content Addressability One very prominent feature of human memory is that it is content addressable. It seems fairly clear that we can access information in memory based on nearly any attribute of the representation we are trying to retrieve.

Of course, some cues are much better than others. An attribute which is shared by a very large number of things we know about is not a very effective retrieval cue, since it does not accurately pick out a particular memory representation. But, several such cues, in conjunction, can do the job. Thus, if we ask a friend who goes out with several women, "Who was that woman I saw you with?" he may not know which one we mean—but if we specify something else about her—say the color of her hair, what she was wearing (in so far as he remembers this at all), where we saw him with her—he will likely be able to hit upon the right one.

It is, of course, possible to implement some kind of content addressability of memory on a standard computer in a variety of different ways. One way is to search sequentially, examining each memory in the system to find the memory or the set of memories which has the particular content specified in the cue. An alternative, somewhat more efficient, scheme involves some form of indexing—keeping a list, for every content a memory might have, of which memories have that content.

Such an indexing scheme can be made to work with error-free probes, but it will break down if there is an error in the specification of the retrieval cue. There are possible ways of recovering from such errors, but they lead to the kind of combinatorial explosions which plague this kind of computer implementation.

But suppose that we imagine that each memory is represented by a unit which has mutually excitatory interactions with units standing for each of its

properties. Then, whenever any property of the memory became active, the memory would tend to be activated, and whenever the memory was activated, all of its contents would tend to become activated. Such a scheme would automatically produce content addressability for us. Though it would not be immune to errors, it would not be devastated by an error in the probe if the remaining properties specified the correct memory.

As described thus far, whenever a property that is a part of a number of different memories is activated, it will tend to activate all of the memories it is in. To keep these other activities from swamping the "correct" memory unit, we simply need to add initial inhibitory connections among the memory units. An additional desirable feature would be mutually inhibitory interactions among mutually incompatible property units. For example, a person cannot both be single and married at the same time, so the units for different marital states would be mutually inhibitory.

McClelland (1981) developed a simulation model that illustrates how a system with these properties would act as a content addressable memory. The model is obviously oversimplified, but it illustrates many of the characteristics of the more complex models that will be considered in later chapters.

Consider the information represented in figure 4.10, which lists a number of people we might meet if we went to live in an unsavory neighborhood, and some of their hypothetical characteristics. A subset of the units needed to represent this information is shown in figure 4.11. In this network, there is an "instance unit" for each of the characters described in figure 4.10, and that unit is linked by mutually excitatory connections to all of the units for the fellow's properties. Note that we have included property units for the names of the characters, as well as units for their other properties.

Now, suppose we wish to retrieve the properties of a particular individual, say Lance. And suppose that we know Lance's name. Then we can probe the network by activating Lance's name unit, and we can see what pattern of activation arises as a result. Assuming that we know of no one else named Lance, we can expect the Lance name unit to be hooked up only to the instance unit for Lance. This will in turn activate the property units for Lance, thereby creating the pattern of activation corresponding to Lance. In effect, we have retrieved a representation of Lance. More will happen than just what we have described so far, but for the moment let us stop here.

Of course, sometimes we may wish to retrieve a name, given other information. In this case, we might start with some of Lance's properties, effectively asking the system, say "Who do you know who is a Shark and in his 20s?" by activating the Shark and 20s units. In this case it turns out that there is a single individual, Ken, who fits the description. So, when we activate these two properties, we will activate the instance unit for Ken, and this in turn will activate his name unit, and fill in his other properties as well.

Graceful Degradation A few of the desirable properties of this kind of model are visible from considering what happens as we vary the set of features we use to probe the memory in an attempt to retrieve a particular individual's name. Any set of features which is sufficient to uniquely characterize a particular item will activate the instance node for that item more strongly than any other in-

The Jets and The Sharks

Name	Gang	Age	Edu	Mar	Occupation
Art	Jets	40's	J.H.	Sing.	Pusher
Al	Jets	30's	J.H.	Mar.	Burglar
Sam	Jets	20's	COL.	Sing.	Bookie
Clyde	Jets	40's	J.H.	Sing.	Bookie
Mike	Jets	30's	J.H.	Sing.	Bookie
Jim	Jets	20's	J.H.	Div.	Burglar
Greg	Jets	20's	H.S.	Mar.	Pusher
John	Jets	20's	J.H.	Mar.	Burglar
Doug	Jets	30's	H.S.	Sing.	Bookie
Lance	Jets	20's	J.H.	Mar.	Burglar
George	Jets	20's	J.H.	Div.	Burglar
Pete	Jets	20's	H.S.	Sing.	Bookie
Fred	Jets	20's	H.S.	Sing.	Pusher
Gene	Jets	20's	COL.	Sing.	Pusher
Ralph	Jets	30's	J.H.	Sing.	Pusher
Phil	Sharks	30's	COL.	Mar.	Pusher
Ike	Sharks	30's	J.H.	Sing.	Bookie
Nick	Sharks	30's	H.S.	Sing.	Pusher
Don	Sharks	30's	COL.	Mar.	Burglar
Ned	Sharks	30's	COL.	Mar.	Bookie
Karl	Sharks	40's	H.S.	Mar.	Bookie
Ken	Sharks	20's	H.S.	Sing.	Burglar
Earl	Sharks	40's	H.S.	Mar.	Burglar
Rick	Sharks	30's	H.S.	Div.	Burglar
Ol	Sharks	30's	COL.	Mar.	Pusher
Neal	Sharks	30's	H.S.	Sing.	Bookie
Dave	Sharks	30's	H.S.	Div.	Pusher

Figure 4.10
Characteristics of a number of individuals belonging to two gangs, the Jets and the Sharks. (From "Retrieving General and Specific Knowledge from Stored Knowledge of Specifics" by J. L. McClelland, 1981, *Proceedings of the Third Annual Conference of the Cognitive Science Society*, Berkeley, CA. Copyright 1981 by J. L. McClelland. Reprinted by permission.)

stance node. A probe which contains misleading features will most strongly activate the node that it matches best. This will clearly be a poorer cue than one which contains no misleading information—but it will still be sufficient to activate the "right answer" more strongly than any other, as long as the introduction of misleading information does not make the probe closer to some other item. In general, though the degree of activation of a particular instance node and of the corresponding name nodes varies in this model as a function of the exact content of the probe, errors in the probe will not be fatal unless they make the probe point to the wrong memory. This kind of model's handling of incomplete or partial probes also requires no special error-recovery scheme to work—it is a natural by-product of the nature of the retrieval mechanism that it is capable of graceful degradation.

These aspects of the behavior of the Jets and Sharks model deserve more detailed consideration than the present space allows. . . . We do, however, have

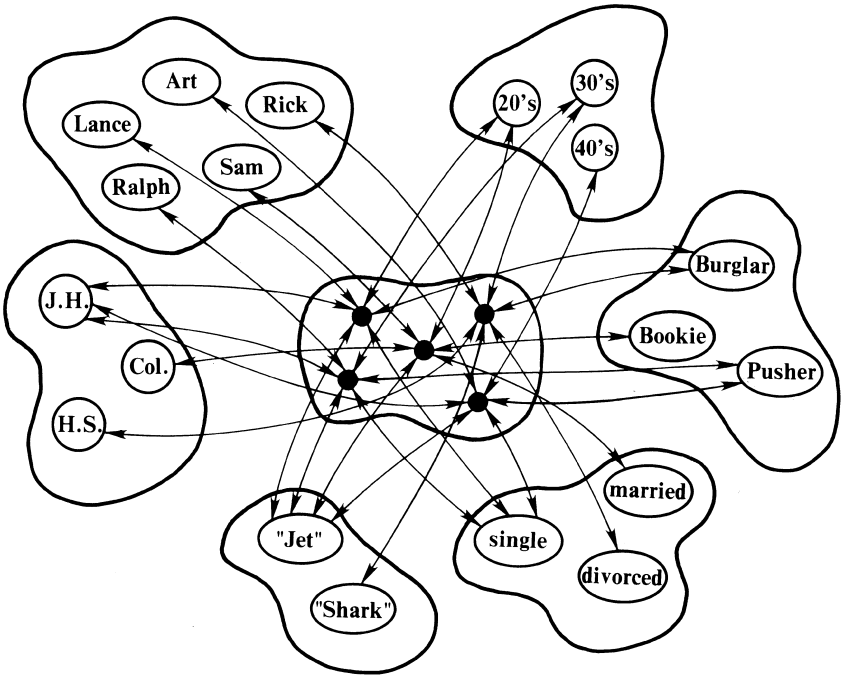


Figure 4.11
Some of the units and interconnections needed to represent the individuals shown in figure 4.10. The units connected with double-headed arrows are mutually excitatory. All the units within the same cloud are mutually inhibitory. (From "Retrieving General and Specific Knowledge from Stored Knowledge of Specifics" by J. L. McClelland, 1981, *Proceedings of the Third Annual Conference of the Cognitive Science Society*, Berkeley, CA. Copyright 1981 by J. L. McClelland. Reprinted by permission.)

more to say about this simple model, for like some of the other models we have already examined, this model exhibits some useful properties which emerge from the interactions of the processing units.

Default Assignment It probably will have occurred to the reader that in many of the situations we have been examining, there will be other activations occurring which may influence the pattern of activation which is retrieved. So, in the case where we retrieved the properties of Lance, those properties, once they become active, can begin to activate the units for other individuals with those same properties. The memory unit for Lance will be in competition with these units and will tend to keep their activation down, but to the extent that they do become active, they will tend to activate their own properties and therefore fill them in. In this way, the model can fill in properties of individuals based on what it knows about other, similar instances.

To illustrate how this might work we have simulated the case in which we do not know that Lance is a Burglar as opposed to a Bookie or a Pusher. It turns out that there are a group of individuals in the set who are very similar to Lance in many respects. When Lance's properties become activated, these other units become partially activated, and they start activating their properties.

Since they all share the same "occupation," they work together to fill in that property for Lance. Of course, there is no reason why this should necessarily be the right answer, but generally speaking, the more similar two things are in respects that we know about, the more likely they are to be similar in respects that we do not, and the model implements this heuristic.

Spontaneous Generalization The model we have been describing has another valuable property as well—it tends to retrieve what is common to those memories which match a retrieval cue which is too general to capture any one memory. Thus, for example, we could probe the system by activating the unit corresponding to membership in the Jets. This unit will partially activate all the instances of the Jets, thereby causing each to send activations to its properties. In this way the model can retrieve the typical values that the members of the Jets have on each dimension—even though there is no one Jet that has these typical values. In the example, 9 of 15 Jets are single, 9 of 15 are in their 20s, and 9 of 15 have only a Junior High School education; when we probe by activating the Jet unit, all three of these properties dominate. The Jets are evenly divided between the three occupations, so each of these units becomes partially activated. Each has a different name, so that each name unit is very weakly activated, nearly cancelling each other out.

In the example just given of spontaneous generalization, it would not be unreasonable to suppose that someone might have explicitly stored a generalization about the members of a gang. The account just given would be an alternative to "explicit storage" of the generalization. It has two advantages, though, over such an account. First, it does not require any special generalization formation mechanism. Second, it can provide us with generalizations on unanticipated lines, on demand. Thus, if we want to know, for example, what people in their 20s with a junior high school education are like, we can probe the model by activating these two units. Since all such people are Jets and Burglars, these two units are strongly activated by the model in this case; two of them are divorced and two are married, so both of these units are partially activated.[1]

The sort of model we are considering, then, is considerably more than a content addressable memory. In addition, it performs default assignment, and it can spontaneously retrieve a general concept of the individuals that match any specifiable probe. These properties must be explicitly implemented as complicated computational extensions of other models of knowledge retrieval, but in PDP models they are natural by-products of the retrieval process itself.

Representation and Learning in PDP Models

In the Jets and Sharks model, we can speak of the model's *active representation* at a particular time, and associate this with the pattern of activation over the units in the system. We can also ask: What is the stored knowledge that gives rise to that pattern of activation? In considering this question, we see immediately an important difference between PDP models and other models of cognitive processes. In most models, knowledge is stored as a static copy of a pattern. Retrieval amounts to finding the pattern in long-term memory and copying it into a buffer or working memory. There is no real difference between the stored representation in long-term memory and the active representation in

working memory. In PDP models, though, this is not the case. In these models, the patterns themselves are not stored. Rather, what is stored is the *connection strengths* between units that allow these patterns to be re-created. In the Jets and Sharks model, there is an instance unit assigned to each individual, but that unit does not contain a copy of the representation of that individual. Instead, it is simply the case that the connections between it and the other units in the system are such that activation of the unit will cause the pattern for the individual to be reinstated on the property units.

This difference between PDP models and conventional models has enormous implications, both for processing and for learning. We have already seen some of the implications for processing. The representation of the knowledge is set up in such a way that the knowledge necessarily influences the course of processing. Using knowledge in processing is no longer a matter of finding the relevant information in memory and bringing it to bear; it is part and parcel of the processing itself.

For learning, the implications are equally profound. For if the knowledge is the strengths of the connections, learning must be a matter of finding the right connection strengths so that the right patterns of activation will be produced under the right circumstances. This is an extremely important property of this class of models, for it opens up the possibility that an information processing mechanism could learn, as a result of tuning its connections, to capture the interdependencies between activations that it is exposed to in the course of processing.

In recent years, there has been quite a lot of interest in learning in cognitive science. Computational approaches to learning fall predominantly into what might be called the "explicit rule formulation" tradition, as represented by the work of Winston (1975), the suggestions of Chomsky, and the ACT* model of J. R. Anderson (1983). All of this work shares the assumption that the goal of learning is to formulate explicit rules (propositions, productions, etc.) which capture powerful generalizations in a succinct way. Fairly powerful mechanisms, usually with considerable innate knowledge about a domain, and/or some starting set of primitive propositional representations, then formulate hypothetical general rules, e.g., by comparing particular cases and formulating explicit generalizations.

The approach that we take in developing PDP models is completely different. First, we do not assume that the goal of learning is the formulation of explicit rules. Rather, we assume it is the acquisition of connection strengths which allow a network of simple units to act *as though* it knew the rules. Second, we do not attribute powerful computational capabilities to the learning mechanism. Rather, we assume very simple connection strength modulation mechanisms which adjust the strength of connections between units based on information locally available at the connection. Our purpose is to give a simple, illustrative example of the connection strength modulation process, and how it can produce networks which exhibit some interesting behavior.

Local versus Distributed Representation Before we turn to an explicit consideration of this issue, we raise a basic question about representation. Once we have achieved the insight that the knowledge is stored in the strengths of the

interconnections between units, a question arises. Is there any reason to assign one unit to each pattern that we wish to learn? Another possibility is that the knowledge about any individual pattern is not stored in the connections of a special unit reserved for that pattern, but is distributed over the connections among a large number of processing units. On this view, the Jets and Sharks model represents a special case in which separate units are reserved for each instance.

Models in which connection information is explicitly thought of as distributed have been proposed by a number of investigators. The units in these collections may themselves correspond to conceptual primitives, or they may have no particular meaning as individuals. In either case, the focus shifts to patterns of activation over these units and to mechanisms whose explicit purpose is to learn the right connection strengths to allow the right patterns of activation to become activated under the right circumstances.

In the rest of this section, we will give a simple example of a PDP model in which the knowledge is distributed. We will first explain how the model would work, given pre-existing connections, and we will then describe how it could come to acquire the right connection strengths through a very simple learning mechanism. A number of models which have taken this distributed approach have been discussed in Hinton and J. A. Anderson's (1981) *Parallel Models of Associative Memory* . We will consider a simple version of a common type of distributed model, a *pattern associator*.

Pattern associators are models in which a pattern of activation over one set of units can cause a pattern of activation over another set of units without any intervening units to stand for either pattern as a whole. Pattern associators would, for example, be capable of associating a pattern of activation on one set of units corresponding to the appearance of an object with a pattern on another set corresponding to the aroma of the object, so that, when an object is presented visually, causing its visual pattern to become active, the model produces the pattern corresponding to its aroma.

How a Pattern Associator Works For purposes of illustration, we present a very simple pattern associator in figure 4.12. In this model, there are four units in each of two pools. The first pool, the A units, will be the pool in which patterns corresponding to the sight of various objects might be represented. The second pool, the B units, will be the pool in which the pattern corresponding to the aroma will be represented. We can pretend that alternative patterns of activation on the A units are produced upon viewing a rose or a grilled steak, and alternative patterns on the B units are produced upon sniffing the same objects. Figure 4.13 shows two pairs of patterns, as well as sets of interconnections necessary to allow the A member of each pair to reproduce the B member.

The details of the behavior of the individual units vary among different versions of pattern associators. For present purposes, we'll assume that the units can take on positive or negative activation values, with 0 representing a kind of neutral intermediate value. The strengths of the interconnections between the units can be positive or negative real numbers.

The effect of an A unit on a B unit is determined by multiplying the activation of the A unit times the strength of its synaptic connection with the B unit.

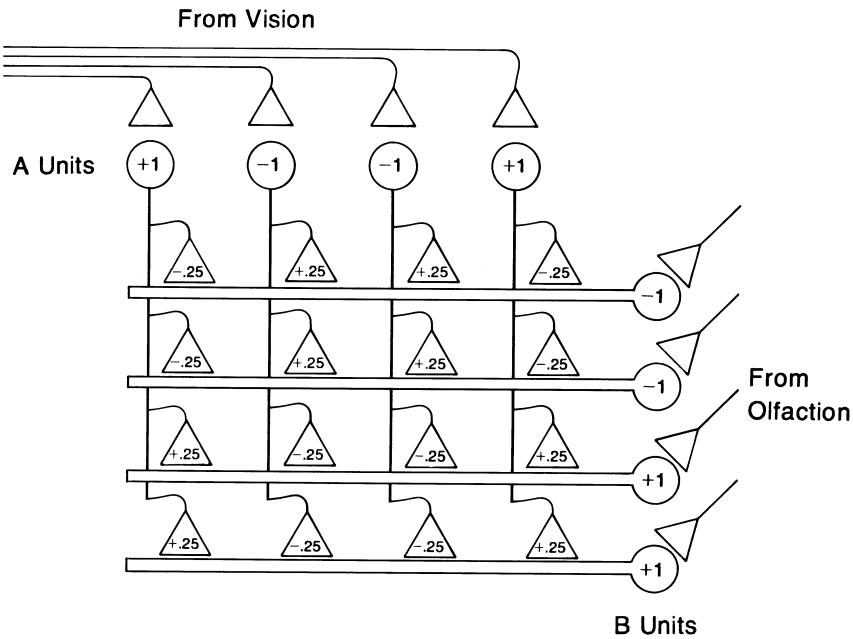


Figure 4.12
A simple pattern associator. The example assumes that patterns of activation in the A units can by produced by the visual system and patterns in the B unit can be produced by the olfactory system. The synaptic connections allow the outputs of the A units to influence the activations of the B units. The synaptic weights linking the A units to the B units were selected so as to allow the pattern of activation shown on the A units to reproduce the pattern of activation shown on the B units without the need for any olfactory input.

	+1	−1	−1	+1
−1	−.25	+.25	+.25	−.25
−1	−.25	+.25	+.25	−.25
+1	+.25	−.25	−.25	+.25
+1	+.25	−.25	−.25	+.25

	−1	+1	−1	+1
−1	+.25	−.25	+.25	−.25
+1	−.25	+.25	−.25	+.25
+1	−.25	+.25	−.25	+.25
−1	+.25	−.25	+.25	−.25

Figure 4.13
Two simple associators represented as matrices. The weights in the first two matrices allow the A pattern shown above the matrix to produce the B pattern shown to the right of it. Note that the weights in the first matrix are the same as those shown in the diagram in figure 4.12.

For example, if the connection from a particular A unit to a particular B unit has a positive sign, when the A unit is excited (activation greater than 0), it will excite the B unit. For this example, we'll simply assume that the activation of each unit is set to the sum of the excitatory and inhibitory effects operating on it. This is one of the simplest possible cases.

Suppose, now, that we have created on the A units the pattern corresponding to the first visual pattern shown in figure 4.13, the rose. How should we arrange the strengths of the interconnections between the A units and the B units to reproduce the pattern corresponding to the aroma of a rose? We simply need to arrange for each A unit to tend to excite each B unit which has a positive activation in the aroma pattern and to inhibit each B unit which has a negative activation in the aroma pattern. It turns out that this goal is achieved by setting the strength of the connection between a given A unit and a given B unit to a value proportional to the product of the activation of the two units. In figure 4.12, the weights on the connections were chosen to allow the A pattern illustrated there to produce the illustrated B pattern according to this principle. The actual strengths of the connections were set to $\pm.25$, rather than ± 1, so that the A pattern will produce the right magnitude, as well as the right sign, for the activations of the units in the B pattern. The same connections are reproduced in matrix form in figure 4.13.

Pattern associators like the one in figure 4.12 have a number of nice properties. One is that they do not require a perfect copy of the input to produce the correct output, though its strength will be weaker in this case. For example, suppose that the associator shown in figure 4.12 were presented with an A pattern of $(1, -1, 0, 1)$. This is the A pattern shown in the figure, with the activation of one of its elements set to 0. The B pattern produced in response will have the activations of all of the B units in the right direction; however, they will be somewhat weaker than they would be, had the complete A pattern been shown. Similar effects are produced if an element of the pattern is distorted—or if the model is damaged, either by removing whole units, or random sets of connections, etc. Thus, their pattern retrieval performance of the model degrades gracefully both under degraded input and under damage.

How a Pattern Associator Learns So far, we have seen how we as model builders can construct the right set of weights to allow one pattern to cause another. The interesting thing, though, is that we do not need to build these interconnection strengths in by hand. Instead, the pattern associator can teach itself the right set of interconnections through experience processing the patterns in conjunction with each other.

A number of different rules for adjusting connection strengths have been proposed. One of the first—and definitely the best known—is due to D. O. Hebb (1949). Hebb's actual proposal was not sufficiently quantitative to build into an explicit model. However, a number of different variants can trace their ancestry back to Hebb. Perhaps the simplest version is:

> When unit A and unit B are simultaneously excited, increase the strength of the connection between them.

A natural extension of this rule to cover the positive and negative activation values allowed in our example is:

> Adjust the strength of the connection between units A and B in proportion to the product of their simultaneous activation.

In this formulation, if the product is positive, the change makes the connection more excitatory, and if the product is negative, the change makes the connection more inhibitory. For simplicity of reference, we will call this the *Hebb rule*, although it is not exactly Hebb's original formulation.

With this simple learning rule, we could train a "blank copy" of the pattern associator shown in figure 4.12 to produce the B pattern for rose when the A pattern is shown, simply by presenting the A and B patterns together and modulating the connection strengths according to the Hebb rule. The size of the change made on every trial would, of course, be a parameter. We generally assume that the changes made on each instance are rather small, and that connection strengths build up gradually. The values shown in figure 4.13, then, would be acquired as a result of a number of experiences with the A and B pattern pair.

It is very important to note that the information needed to use the Hebb rule to determine the value each connection should have is *locally available* at the connection. All a given connection needs to consider is the activation of the units on both sides of it. Thus, it would be possible to actually implement such a connection modulation scheme locally, in each connection, without requiring any programmer to reach into each connection and set it to just the right value.

It turns out that the Hebb rule as stated here has some serious limitations, and, to our knowledge, no theorists continue to use it in this simple form. More sophisticated connection modulation schemes have been proposed by other workers; most important among these are the delta rule; the competitive learning rule; and the rules for learning in stochastic parallel models. All of these learning rules have the property that they adjust the strengths of connections between units on the basis of information that can be assumed to be locally available to the unit. Learning, then, in all of these cases, amounts to a very simple process that can be implemented locally at each connection without the need for any overall supervision. Thus, models which incorporate these learning rules train themselves to have the right interconnections in the course of processing the members of an ensemble of patterns.

Learning Multiple Patterns in the Same Set of Interconnections Up to now, we have considered how we might teach our pattern associator to associate the visual pattern for one object with a pattern for the aroma of the same object. Obviously, different patterns of interconnections between the A and B units are appropriate for causing the visual pattern for a different object to give rise to the pattern for its aroma. The same principles apply, however, and if we presented our pattern associator with the A and B patterns for steak, it would learn the right set of interconnections for that case instead (these are shown in figure 4.13). In fact, it turns out that we can actually teach the same pattern associator a number of different associations. The matrix representing the set of interconnections that would be learned if we taught the same pattern associator

$$\begin{bmatrix} - & + & + & - \\ - & + & + & - \\ + & - & - & + \\ + & - & - & + \end{bmatrix} + \begin{bmatrix} + & - & + & - \\ - & + & - & + \\ - & + & - & + \\ + & - & + & - \end{bmatrix} = \begin{bmatrix} & & ++ & -- \\ -- & ++ & & \\ & & -- & ++ \\ ++ & -- & & \end{bmatrix}$$

Figure 4.14
The weights in the third matrix allow either A pattern shown in figure 4.13 to recreate the corresponding B pattern. Each weight in this case is equal to the sum of the weight for the A pattern and the weight for the B pattern, as illustrated.

both the rose association and the steak association is shown in figure 4.14. The reader can verify this by adding the two matrices for the individual patterns together. The reader can also verify that this set of connections will allow the rose A pattern to produce the rose B pattern, and the steak A pattern to produce the steak B pattern: when either input pattern is presented, the correct corresponding output is produced.

The examples used here have the property that the two different visual patterns are completely uncorrelated with each other. This being the case, the rose pattern produces no effect when the interconnections for the steak have been established, and the steak pattern produces no effect when the interconnections for the rose association are in effect. For this reason, it is possible to add together the pattern of interconnections for the rose association and the pattern for the steak association, and still be able to associate the sight of the steak with the smell of a steak and the sight of a rose with the smell of a rose. The two sets of interconnections do not interact at all.

One of the limitations of the Hebbian learning rule is that it can learn the connection strengths appropriate to an entire ensemble of patterns only when all the patterns are completely uncorrelated. This restriction does not, however, apply to pattern associators which use more sophisticated learning schemes.

Attractive Properties of Pattern Associator Models Pattern associator models have the property that uncorrelated patterns do not interact with each other, but more similar ones do. Thus, to the extent that a new pattern of activation on the A units is similar to one of the old ones, it will tend to have similar effects. Furthermore, if we assume that learning the interconnections occurs in small increments, similar patterns will essentially reinforce the strengths of the links they share in common with other patterns. Thus, if we present the same pair of patterns over and over, but each time we add a little random noise to each element of each member of the pair, the system will automatically learn to associate the central tendency of the two patterns and will learn to ignore the noise. What will be stored will be an average of the similar patterns with the slight variations removed. On the other hand, when we present the system with completely uncorrelated patterns, they will not interact with each other in this way. Thus, the same pool of units can extract the central tendency of each of a number of pairs of unrelated patterns.

Extracting the Structure of an Ensemble of Patterns The fact that similar patterns tend to produce similar effects allows distributed models to exhibit a kind of

spontaneous generalization, extending behavior appropriate for one pattern to other similar patterns. This property is shared by other PDP models, such as the word perception model and the Jets and Sharks model described above; the main difference here is in the existence of simple, local, learning mechanisms that can allow the acquisition of the connection strengths needed to produce these generalizations through experience with members of the ensemble of patterns. Distributed models have another interesting property as well: If there are regularities in the correspondences between pairs of patterns, the model will naturally extract these regularities. This property allows distributed models to acquire patterns of interconnections that lead them to behave in ways we ordinarily take as evidence for the use of linguistic rules.

We describe one such model very briefly. The model is a mechanism that learns how to construct the past tenses of words from their root forms through repeated presentations of examples of root forms paired with the corresponding past-tense form. The model consists of two pools of units. In one pool, patterns of activation representing the phonological structure of the root form of the verb can be represented, and, in the other, patterns representing the phonological structure of the past tense can be represented. The goal of the model is simply to learn the right connection strengths between the root units and the past-tense units, so that whenever the root form of a verb is presented the model will construct the corresponding past-tense form. The model is trained by presenting the root form of the verb as a pattern of activation over the root units, and then using a simple, local, learning rule to adjust the connection strengths so that this root form will tend to produce the correct pattern of activation over the past-tense units. The model is tested by simply presenting the root form as a pattern of activation over the root units and examining the pattern of activation produced over the past-tense units.

The model is trained initially with a small number of verbs children learn early in the acquisition process. At this point in learning, it can only produce appropriate outputs for inputs that it has explicitly been shown. But as it learns more and more verbs, it exhibits two interesting behaviors. First, it produces the standard *ed* past tense when tested with pseudo-verbs or verbs it has never seen. Second, it "overregularizes" the past tense of irregular words it previously completed correctly. Often, the model will blend the irregular past tense of the word with the regular *ed* ending, and produce errors like *CAMED* as the past of *COME*. These phenomena mirror those observed in the early phases of acquisition of control over past tenses in young children.

The generativity of the child's responses—the creation of regular past tenses of new verbs and the overregularization of the irregular verbs—has been taken as strong evidence that the child has induced the rule which states that the regular correspondence for the past tense in English is to add a final *ed* (Berko, 1958). On the evidence of its performance, then, the model can be said to have acquired the rule. However, no special rule-induction mechanism is used, and no special language-acquisition device is required. The model learns to behave in accordance with the rule, not by explicitly noting that most words take *ed* in the past tense in English and storing this rule away explicitly, but simply by building up a set of connections in a pattern associator through a long series of simple learning experiences. The same mechanisms of parallel distributed

processing and connection modification which are used in a number of domains serve, in this case, to produce implicit knowledge tantamount to a linguistic rule. The model also provides a fairly detailed account of a number of the specific aspects of the error patterns children make in learning the rule. In this sense, it provides a richer and more detailed description of the acquisition process than any that falls out naturally from the assumption that the child is building up a repertoire of explicit but inaccessible rules.

There is a lot more to be said about distributed models of learning, about their strengths and their weaknesses, than we have space for in this brief consideration. For now we hope mainly to have suggested that they provide dramatically different accounts of learning and acquisition than are offered by traditional models of these processes. We saw in earlier sections of this chapter that performance in accordance with rules can emerge from the interactions of simple, interconnected units. Now we can see how the aquisition of performance that conforms to linguistic rules can emerge from a simple, local, connection strength modulation process.

We have seen what the properties of PDP models are in informal terms, and we have seen how these properties operate to make the models do many of the kinds of things that they do. We now wish to describe some of the major sources of inspiration for the PDP approach.

Origins of Parallel Distributed Processing

The ideas behind the PDP approach have a history that stretches back indefinitely. In this section, we mention briefly some of the people who have thought in these terms, particularly those whose work has had an impact on our own thinking. This section should not been seen as an authoritative review of the history, but only as a description of our own sources of inspiration.

Some of the earliest roots of the PDP approach can be found in the work of the unique neurologists, Jackson (1869/1958) and Luria (1966). Jackson was a forceful and persuasive critic of the simplistic localizationist doctrines of late nineteenth century neurology, and he argued convincingly for distributed, multilevel conceptions of processing systems. Luria, the Russian psychologist and neurologist, put forward the notion of the *dynamic functional system*. On this view, every behavioral or cognitive process resulted from the coordination of a large number of different components, each roughly localized in different regions of the brain, but all working together in dynamic interaction. Neither Hughlings-Jackson nor Luria is noted for the clarity of his views, but we have seen in their ideas a rough characterization of the kind of parallel distributed processing system we envision.

Two other contributors to the deep background of PDP were Hebb (1949) and Lashley (1950). We already have noted Hebb's contribution of the Hebb rule of synaptic modification; he also introduced the concept of cell assemblies—a concrete example of a limited form of distributed processing—and discussed the idea of reverberation of activation within neural networks. Hebb's ideas were cast more in the form of speculations about neural functioning than in the form of concrete processing models, but his thinking captures some of the flavor of parallel distributed processing mechanisms. Lashley's contribution was

to insist upon the idea of distributed representation. Lashley may have been too radical and too vague, and his doctrine of equipotentiality of broad regions of cortex clearly overstated the case. Yet many of his insights into the difficulties of storing the "engram" locally in the brain are telling, and he seemed to capture quite precisely the essence of distributed representation in insisting that "there are no special cells reserved for special memories" (Lashley, 1950, p. 500).

In the 1950s, there were two major figures whose ideas have contributed to the development of our approach. One was Rosenblatt (1959, 1962) and the other was Selfridge (1955). In his *Principles of Neurodynamics* (1962), Rosenblatt articulated clearly the promise of a neurally inspired approach to computation, and he developed the *perceptron convergence procedure*, an important advance over the Hebb rule for changing synaptic connections. Rosenblatt's work was very controversial at the time, and the specific models he proposed were not up to all the hopes he had for them. But his vision of the human information processing system as a dynamic, interactive, self-organizing system lies at the core of the PDP approach. Selfridge's contribution was his insistence on the importance of interactive processing, and the development of *Pandemonium*, an explicitly computational example of a dynamic, interactive mechanism applied to computational problems in perception.

In the late 60s and early 70s, serial processing and the von Neumann computer dominated both psychology and artificial intelligence, but there were a number of researchers who proposed neural mechanisms which capture much of the flavor of PDP models. Among these figures, the most influential in our work have been J. A. Anderson, Grossberg, and Longuet-Higgins. Grossberg's mathematical analysis of the properties of neural networks led him to many insights we have only come to appreciate through extensive experience with computer simulation, and he deserves credit for seeing the relevance of neurally inspired mechanisms in many areas of perception and memory well before the field was ready for these kinds of ideas (Grossberg, 1978). Grossberg (1976) was also one of the first to analyze certain properties of the competitive learning mechanism. Anderson's work differs from Grossberg's in insisting upon distributed representation, and in showing the relevance of neurally inspired models for theories of concept learning (Anderson, 1973, 1977); work on distributed memory and amnesia owes a great deal to Anderson's inspiration. Anderson's work also played a crucial role in the formulation of the *cascade* model (McClelland, 1979), a step away from serial processing down the road to PDP. Longuet-Higgins and his group at Edinburgh were also pursuing distributed memory models during the same period, and David Willshaw, a member of the Edinburgh group, provided some very elegant mathematical analyses of the properties of various distributed representation schemes (Willshaw, 1981). His insights provide one of the sources of the idea of coarse coding. Many of the contributions of Anderson, Willshaw, and others distributed modelers may be found in Hinton and Anderson (1981). Others who have made important contributions to learning in PDP models include Amari (1977), Bienenstock, Cooper, and Munro (1982), Fukushima (1975), Kohonen (1977, 1984), and von der Malsburg (1973).

Toward the middle of the 1970s, the idea of parallel processing began to have something of a renaissance in computational circles. We have already mentioned the Marr and Poggio (1976) model of stereoscopic depth perception. Another model from this period, the *HEARSAY* model of speech understanding, played a prominent role in the development of our thinking. Unfortunately, HEARSAY's computational architecture was too demanding for the available computational resources, and so the model was not a computational success. But its basically parallel, interactive character inspired the interactive model of reading (Rumelhart, 1977), and the interactive activation model of word recognition (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982).

The ideas represented in the interactive activation model had other precursors as well. Morton's *logogen* model (Morton, 1969) was one of the first models to capture concretely the principle of interaction of different sources of information, and Marslen-Wilson (e.g., Marslen-Wilson & Welsh, 1978) provided important empirical demonstrations of interaction between different levels of language processing. Levin's (1976) *Proteus* model demonstrated the virtues of activation-competition mechanisms, and Glushko (1979) helped us see how conspiracies of partial activations could account for certain aspects of apparently rule-guided behavior.

Our work also owes a great deal to a number of colleagues who have been working on related ideas in recent years. Feldman and Ballard (1982) laid out many of the computational principles of the PDP approach (under the name of *connectionism*), and stressed the biological implausibility of most of the prevailing computational models in artificial intelligence. Hofstadter (1979, 1985) deserves credit for stressing the existence of a subcognitive—what we call microstructural—level, and pointing out how important it can be to delve into the microstructure to gain insight. A sand dune, he has said, is not a grain of sand. Others have contributed crucial technical insights. Sutton and Barto (1981) provided an insightful analysis of the connection modification scheme we call the *delta rule* and illustrated the power of the rule to account for some of the subtler properties of classical conditioning. And Hopfield's (1982) contribution of the idea that network models can be seen as seeking minima in energy landscapes played a prominent role in the development of the Boltzmann machine and in the crystallization of ideas on harmony theory and schemata.

The power of parallel distributed processing is becoming more and more apparent, and many others have recently joined in the exploration of the capabilities of these mechanisms. We hope this chapter represents the nature of the enterprise we are all involved in, and that it does justice to the potential of the PDP approach.

Acknowledgments

This research was supported by Contract N00014-79-C-0323, NR667-437 with the Personnel and Training Research Programs of the Office of Naval Research, by grants from the System Development Foundation, and by a NIMH Career Development Award (MH00385) to the first author.

Note

1. In this and all other cases, there is a tendency for the pattern of activation to be influenced by partially activated, near neighbors, which do not quite match the probe. Thus, in this case, there is a Jet Al, who is a Married Burglar. The unit for Al gets slightly activated, giving Married a slight edge over Divorced in the simulation.

References

Amari, S. A. (1977). A mathematical approach to neural systems. In J. Metzler (Ed.), *Systems neuroscience* (pp. 67–117). New York: Academic Press.

Anderson, J. R. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review, 80*, 417–438.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Berko, J. (1958). The child's learning of English morphology. *Word, 14*, 150–177.

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron activity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience, 2*, 32–48.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science, 6*, 205–254.

Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics, 20*, 121–136.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 674–691.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: Part I. Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23*, 121–134.

Grossberg, S. (1978). A theory of visual coding, memory, and development. In E. L. J. Leeuwenberg & H. F. J. M. Buffart (Eds.), *Formal theories of visual perception*. New York: Wiley.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hinton, G. E., & Anderson, J. A. (Eds.). (1981). *Parallel models of associative memory*. Hillsdale, NJ: Erlbaum.

Hinton, G. E., & Anderson, J. A. (Eds.). (1981). *Parallel models of associative memory*. Hillsdale, NJ: Erlbaum.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An eternal golden braid*. New York: Basic Books.

Hofstadter, D. R. (1985). *Metamagical themas*. New York: Basic Books.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings from the National Academy of Sciences, USA, 81*, 6871–6874.

Isenberg, D., Walker, E. C. T., Ryder, J. M., & Schweikert, J. (1980, November). *A top-down effect on the identification of function words*. Paper presented at the Acoustical Society of America, Los Angeles.

Jackson, J. H. (1869/1958). On localization. In *Selected writings* (Vol. 2). New York: Basic Books. (Original work published in 1869).

Kohonen, T. (1977). *Associative memory: A system theoretical approach*. New York: Springer.

Kohonen, T. (1984). *Self-organization and associative memory*. Berlin: Springer-Verlag.

Lashley, K. S. (1950). In search of the engram. In *Society of Experimental Biology Symposium No. 4: Psychological mechanisms in animal behavior* (pp. 478–505). London: Cambridge University Press.

Levin, J. A. (1976). *Proteus: An activation framework for cognitive process models* (Tech. Rep. No. ISI/WP-2). Marina de Rey, CA: University of Southern California, Information Sciences Institute.

Lindsay, P. H., & Norman, D. A. (1972). *Human information processing: An introduction to psychology*, New York: Academic Press.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29–63.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London, Series B, 204*, 301–328.

McClelland, J. L. (1979). On the time-relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*, 287–330.

McClelland, J. L. (1981). Retrieving general and specific information from stored knowledge of specifics. *Proceedings of the Third Annual Meeting of the Cognitive Science Society*, 170–172.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375–407.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). New York: McGraw-Hill.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*, 165–178.

Norman, D. A., & Bobrow, D. G. (1976). On the role of active memory processes in perception and cognition. In C. N. Cofer (Ed.), *The structure of human memory* (pp. 114–132). San Francisco: Freeman.

Pillsbury, W. B. (1897). A study in appreception. *American Journal of Psychology, 8*, 315–393.

Rosenblatt, F. (1959). Two theorems of statistical separability in the perceptron. In *Mechanisation of thought processes: Proceedings of a symposium held at the National Physics Laboratory, November 1958. Vol. 1* (pp. 421–456). London: HM Stationary Office.

Rumelhart, D. E. (1975). Notes on a schema for stories. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding* (pp. 211–236). New York: Academic Press.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention & Performance VI*. Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89*, 60–94.

Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science, 6*, 1–36.

Schank, R. C. (1973). Identification of conceptualizations underlying natural language. In R. C. Schank & K. M. Colby (Eds.), *Computer models of thought and langauge* (pp. 187–247). San Francisco: Freeman.

Schank, R. C. (1976). The role of memory in language processing. In C. N. Cofer (Ed.), *The structure of human memory* (pp. 162–189). San Francisco: Freeman.

Selfridge, O. G. (1955). Pattern recognition in modern computers. *Proceedings of the Western Joint Computer Conference*.

Smith, P. T., & Baker, R. G. (1976). The influence of English spelling patterns on pronounciation. *Journal of Verbal Learning and Verbal Behaviour, 15*, 267–286.

Spoehr, K., & Smith, E. (1975). The role of orthographic and phonotactic rules in perceiving letter patterns. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 21–34.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review, 88*, 135–170.

Venesky, R. L. (1970). *The structure of Ensligh orthography*. The Hague: Mouton.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik, 14*, 85–100.

Warren, R. M. (1970). Perceptual restortation of missing speech sounds. *Science, 167*, 393–395.

Willshaw, D. J. (1981). Holography, associative memory, and inductive generalization. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 83–104). Hillsdale, NJ: Erlbaum.

Winston, P. H. (1975). Learning structural descriptions from examples. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 157–209). New York: McGraw-Hill.

PART III

Objections

Chapter 5

Minds, Brains, and Programs

John R. Searle

What psychological and philosophical significance should we attach to recent efforts at computer simulations of human cognitive capacities? In answering this question, I find it useful to distinguish what I will call "strong" AI from "weak" or "cautious" AI (Artificial Intelligence). According to weak AI, the principal value of the computer in the study of the mind is that it gives us a very powerful tool. For example, it enables us to formulate and test hypotheses in a more rigorous and precise fashion. But according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really *is* a mind, in the sense that computers given the right programs can be literally said to *understand* and have other cognitive states. In strong AI, because the programmed computer has cognitive states, the programs are not mere tools that enable us to test psychological explanations; rather, the programs are themselves the explanations.

I have no objection to the claims of weak AI, at least as far as this article is concerned. My discussion here will be directed at the claims I have defined as those of strong AI, specifically the claim that the appropriately programmed computer literally has cognitive states and that the programs thereby explain human cognition. When I hereafter refer to AI, I have in mind the strong version, as expressed by these two claims.

I will consider the work of Roger Schank and his colleagues at Yale (Schank and Abelson, 1977), because I am more familiar with it than I am with any other similar claims, and because it provides a very clear example of the sort of work I wish to examine. But nothing that follows depends upon the details of Schank's programs. The same arguments would apply to Winograd's SHRDLU (Winograd, 1973), Weizenbaum's ELIZA (Weizenbaum, 1965), and indeed any Turing machine simulation of human mental phenomena.

Very briefly, and leaving out the various details, one can describe Schank's program as follows: the aim of the program is to simulate the human ability to understand stories. It is characteristic of human beings' story-understanding capacity that they can answer questions about the story even though the information that they give was never explicitly stated in the story. Thus, for example, suppose you are given the following story: "A man went into a restaurant and ordered a hamburger. When the hamburger arrived it was burned to a crisp, and the man stormed out of the restaurant angrily, without paying for the hamburger or leaving a tip." Now, if you are asked "Did the man eat the

From *The Behavioral and Brain Sciences* 3 (1980): 140–152. Reprinted with permission.

hamburger?" you will presumably answer, "No, he did not." Similarly, if you are given the following story: "A man went into a restaurant and ordered a hamburger; when the hamburger came he was very pleased with it; and as he left the restaurant he gave the waitress a large tip before paying his bill," and you are asked the question, "Did the man eat the hamburger?" you will presumably answer, "Yes, he ate the hamburger." Now Schank's machines can similarly answer questions about restaurants in this fashion. To do this, they have a "representation" of the sort of information that human beings have about restaurants, which enables them to answer such questions as those above, given these sorts of stories. When the machine is given the story and then asked the question, the machine will print out answers of the sort that we would expect human beings to give if told similar stories. Partisans of strong AI claim that in this question and answer sequence the machine is not only simulating a human ability but also

1. that the machine can literally be said to *understand* the story and provide the answers to questions, and
2. that what the machine and its program do *explains* the human ability to understand the story and answer questions about it.

Both claims seem to me to be totally unsupported by Schank's[1] work, as I will attempt to show in what follows.

One way to test any theory of the mind is to ask oneself what it would be like if my mind actually worked on the principles that the theory says all minds work on. Let us apply this test to the Schank program with the following *Gedankenexperiment*. Suppose that I'm locked in a room and given a large batch of Chinese writing. Suppose furthermore (as is indeed the case) that I know no Chinese, either written or spoken, and that I'm not even confident that I could recognize Chinese writing as Chinese writing distinct from, say, Japanese writing or meaningless squiggles. To me, Chinese writing is just so many meaningless squiggles. Now suppose further that after this first batch of Chinese writing I am given a second batch of Chinese script together with a set of rules for correlating the second batch with the first batch. The rules are in English, and I understand these rules as well as any other native speaker of English. They enable me to correlate one set of formal symbols with another set of formal symbols, and all that "formal" means here is that I can identify the symbols entirely by their shapes. Now suppose also that I am given a third batch of Chinese symbols together with some instructions, again in English, that enable me to correlate elements of this third batch with the first two batches, and these rules instruct me how to give back certain Chinese symbols with certain sorts of shapes in response to certain sorts of shapes given me in the third batch. Unknown to me, the people who are giving me all of these symbols call the first batch "a script," they call the second batch a "story," and they call the third batch "questions." Furthermore, they call the symbols I give them back in response to the third batch "answers to the questions," and the set of rules in English that they gave me, they call "the program." Now just to complicate the story a little, imagine that these people also give me stories in English, which I understand, and they then ask me questions in English about these stories, and I give them back answers in English. Suppose also that after a while I get so

good at following the instructions for manipulating the Chinese symbols and the programmers get so good at writing the programs that from the external point of view—that is, from the point of view of somebody outside the room in which I am locked—my answers to the questions are absolutely indistinguishable from those of native Chinese speakers. Nobody just looking at my answers can tell that I don't speak a word of Chinese. Let us also suppose that my answers to the English questions are, as they no doubt would be, indistinguishable from those of other native English speakers, for the simple reason that I am a native English speaker. From the external point of view—from the point of view of someone reading my "answers"—the answers to the Chinese questions and the English questions are equally good. But in the Chinese case, unlike the English case, I produce the answers by manipulating uninterpreted formal symbols. As far as the Chinese is concerned, I simply behave like a computer; I perform computational operations on formally specified elements. For the purposes of the Chinese, I am simply an instantiation of the computer program.

Now the claims made by strong AI are that the programmed computer understands the stories and that the program in some sense explains human understanding. But we are now in a position to examine these claims in light of our thought experiment.

1. As regards the first claim, it seems to me quite obvious in the example that I do not understand a word of the Chinese stories. I have inputs and outputs that are indistinguishable from those of the native Chinese speaker, and I can have any formal program you like, but I still understand nothing. For the same reasons, Schank's computer understands nothing of any stories, whether in Chinese, English, or whatever, since in the Chinese case the computer is me, and in cases where the computer is not me, the computer has nothing more than I have in the case where I understand nothing.

2. As regards the second claim, that the program explains human understanding, we can see that the computer and its program do not provide sufficient conditions of understanding since the computer and the program are functioning, and there is no understanding. But does it even provide a necessary condition or a significant contribution to understanding? One of the claims made by the supporters of strong AI is that when I understand a story in English, what I am doing is exactly the same—or perhaps more of the same—as what I was doing in manipulating the Chinese symbols. It is simply more formal symbol manipulation that distinguishes the case in English, where I do understand, from the case in Chinese, where I don't. I have not demonstrated that this claim is false, but it would certainly appear an incredible claim in the example. Such plausibility as the claim has derives from the supposition that we can construct a program that will have the same inputs and outputs as native speakers, and in addition we assume that speakers have some level of description where they are also instantiations of a program. On the basis of these two assumptions we assume that even if Schank's program isn't the whole story about understanding, it may be part of the story. Well, I suppose that is an empirical possibility, but not the slightest reason has so far been given to believe that it is true, since what is suggested—though certainly not demonstrated—by the example is that the computer program is simply

irrelevant to my understanding of the story. In the Chinese case I have everything that artificial intelligence can put into me by way of a program, and I understand nothing; in the English case I understand everything, and there is so far no reason at all to suppose that my understanding has anything to do with computer programs, that is, with computational operations on purely formally specified elements. As long as the program is defined in terms of computational operations on purely formally defined elements, what the example suggests is that these by themselves have no interesting connection with understanding. They are certainly not sufficient conditions, and not the slightest reason has been given to suppose that they are necessary conditions or even that they make a significant contribution to understanding. Notice that the force of the argument is not simply that different machines can have the same input and output while operating on different formal principles—that is not the point at all. Rather, whatever purely formal principles you put into the computer, they will not be sufficient for understanding, since a human will be able to follow the formal principles without understanding anything. No reason whatever has been offered to suppose that such principles are necessary or even contributory, since no reason has been given to suppose that when I understand English I am operating with any formal program at all.

Well, then, what is it that I have in the case of the English sentences that I do not have in the case of the Chinese sentences? The obvious answer is that I know what the former mean, while I haven't the faintest idea what the latter mean. But in what does this consist and why couldn't we give it to a machine, whatever it is? I will return to this question later, but first I want to continue with the example.

I have had the occasion to present this example to several workers in artificial intelligence, and, interestingly, they do not seem to agree on what the proper reply to it is. I get a surprising variety of replies, and in what follows I will consider the most common of these (specified along with their geographic origins).

But first I want to block some common misunderstandings about "understanding": in many of these discussions one finds a lot of fancy footwork about the word "understanding." My critics point out that there are many different degrees of understanding; that "understanding" is not a simple two-place predicate; that there are even different kinds and levels of understanding, and often the law of excluded middle doesn't even apply in a straightforward way to statements of the form "x understands y"; that in many cases it is a matter for decision and not a simple matter of fact whether x understands y; and so on. To all of these points I want to say: of course, of course. But they have nothing to do with the points at issue. There are clear cases in which "understanding" literally applies and clear cases in which it does not apply; and these two sorts of cases are all I need for this argument.[2] I understand stories in English; to a lesser degree I can understand stories in French; to a still lesser degree, stories in German; and in Chinese, not at all. My car and my adding machine, on the other hand, understand nothing: they are not in that line of business. We often attribute "understanding" and other cognitive predicates by metaphor and analogy to cars, adding machines, and other artifacts, but nothing is proved by such attributions. We say, "The door *knows* when to open because of its photo-

electric cell," "The adding machine *knows how* (*understands how, is able*) to do addition and subtraction but not division," and "The thermostat *perceives* changes in the temperature." The reason we make these attributions is quite interesting, and it has to do with the fact that in artifacts we extend our own intentionality;[3] our tools are extensions of our purposes, and so we find it natural to make metaphorical attributions of intentionality to them; but I take it no philosophical ice is cut by such examples. The sense in which an automatic door "understands instructions" from its photoelectric cell is not at all the sense in which I understand English. If the sense in which Schank's programmed computers understand stories is supposed to be the metaphorical sense in which the door understands, and not the sense in which I understand English, the issue would not be worth discussing. But Newell and Simon (1963) write that the kind of cognition they claim for computers is exactly the same as for human beings. I like the straightforwardness of this claim, and it is the sort of claim I will be considering. I will argue that in the literal sense the programmed computer understands what the car and the adding machine understand, namely, exactly nothing. The computer understanding is not just (like my understanding of German) partial or incomplete; it is zero.

Now to the replies.

5.1 The Systems Reply (Berkeley)

"While it is true that the individual person who is locked in the room does not understand the story, the fact is that he is merely part of a whole system, and the system does understand the story. The person has a large ledger in front of him in which are written the rules, he has a lot of scratch paper and pencils for doing calculations, he has 'data banks' of sets of Chinese symbols. Now, understanding is not being ascribed to the mere individual; rather it is being ascribed to this whole system of which he is a part."

My response to the systems theory is quite simple: let the individual internalize all of these elements of the system. He memorizes the rules in the ledger and the data banks of Chinese symbols, and he does all the calculations in his head. The individual then incorporates the entire system. There isn't anything at all to the system that he does not encompass. We can even get rid of the room and suppose he works outdoors. All the same, he understands nothing of the Chinese, and a fortiori neither does the system, because there isn't anything in the system that isn't in him. If he doesn't understand, then there is no way the system could understand because the system is just a part of him.

Actually I feel somewhat embarrassed to give even this answer to the systems theory because the theory seems to me so unplausible to start with. The idea is that while a person doesn't understand Chinese, somehow the *conjunction* of that person and bits of paper might understand Chinese. It is not easy for me to imagine how someone who was not in the grip of an ideology would find the idea at all plausible. Still, I think many people who are committed to the ideology of strong AI will in the end be inclined to say something very much like this; so let us pursue it a bit further. According to one version of this view, while the man in the internalized systems example doesn't understand Chinese in the sense that a native Chinese speaker does (because, for example,

he doesn't know that the story refers to restaurants and hamburgers, etc.), still "the man as a formal symbol manipulation system" *really does understand Chinese*. The subsystem of the man that is the formal symbol manipulation system for Chinese should not be confused with the subsystem for English.

So there are really two subsystems in the man; one understands English, the other Chinese, and "it's just that the two systems have little to do with each other." But, I want to reply, not only do they have little to do with each other, they are not even remotely alike. The subsystem that understands English (assuming we allow ourselves to talk in this jargon of "subsystems" for a moment) knows that the stories are about restaurants and eating hamburgers, he knows that he is being asked questions about restaurants and that he is answering questions as best he can by making various inferences from the content of the story, and so on. But the Chinese system knows none of this. Whereas the English subsystem knows that "hamburgers" refers to hamburgers, the Chinese subsystem knows only that "squiggle squiggle" is followed by "squoggle squoggle." All he knows is that various formal symbols are being introduced at one end and manipulated according to rules written in English, and other symbols are going out at the other end. The whole point of the original example was to argue that such symbol manipulation by itself couldn't be sufficient for understanding Chinese in any literal sense because the man could write "squoggle squoggle" after "squiggle squiggle" without understanding anything in Chinese. And it doesn't meet that argument to postulate subsystems within the man, because the subsystems are no better off than the man was in the first place; they still don't have anything even remotely like what the English-speaking man (or subsystem) has. Indeed, in the case as described, the Chinese subsystem is simply a part of the English subsystem, a part that engages in meaningless symbol manipulation according to rules in English.

Let us ask ourselves what is supposed to motivate the systems reply in the first place; that is, what *independent* grounds are there supposed to be for saying that the agent must have a subsystem within him that literally understands stories in Chinese? As far as I can tell the only grounds are that in the example I have the same input and output as native Chinese speakers and a program that goes from one to the other. But the whole point of the example has been to try to show that that couldn't be sufficient for understanding, in the sense in which I understand stories in English, because a person, and hence the set of systems that go to make up a person, could have the right combination of input, output, and program and still not understand anything in the relevant literal sense in which I understand English. The only motivation for saying there *must* be a subsystem in me that understands Chinese is that I have a program and I can pass the Turing test; I can fool native Chinese speakers. But precisely one of the points at issue is the adequacy of the Turing test. The example shows that there could be two "systems," both of which pass the Turing test, but only one of which understands; and it is no argument against this point to say that since they both pass the Turing test they must both understand, since this claim fails to meet the argument that the system in me that understands English has a great deal more than the system that merely processes Chinese. In short, the

systems reply simply begs the question by insisting without argument that the system must understand Chinese.

Furthermore, the systems reply would appear to lead to consequences that are independently absurd. If we are to conclude that there must be cognition in me on the grounds that I have a certain sort of input and output and a program in between, then it looks like all sorts of noncognitive subsystems are going to turn out to be cognitive. For example, there is a level of description at which my stomach does information processing, and it instantiates any number of computer programs, but I take it we do not want to say that it has any understanding (cf. Pylyshyn, 1980). But if we accept the systems reply, then it is hard to see how we avoid saying that stomach, heart, liver, and so on, are all understanding subsystems, since there is no principled way to distinguish the motivation for saying the Chinese subsystem understands from saying that the stomach understands. It is, by the way, not an answer to this point to say that the Chinese system has information as input and output and the stomach has food and food products as input and output, since from the point of view of the agent, from my point of view, there is no information in either the food or the Chinese—the Chinese is just so many meaningless squiggles. The information in the Chinese case is solely in the eyes of the programmers and the interpreters, and there is nothing to prevent them from treating the input and output of my digestive organs as information if they so desire.

This last point bears on some independent problems in strong AI, and it is worth digressing for a moment to explain it. If strong AI is to be a branch of psychology, then it must be able to distinguish those systems that are genuinely mental from those that are not. It must be able to distinguish the principles on which the mind works from those on which nonmental systems work; otherwise it will offer us no explanations of what is specifically mental about the mental. And the mental–nonmental distinction cannot be just in the eye of the beholder but it must be intrinsic to the systems; otherwise it would be up to any beholder to treat people as nonmental and, for example, hurricanes as mental if he likes. But quite often in the AI literature the distinction is blurred in ways that would in the long run prove disastrous to the claim that AI is a cognitive inquiry. McCarthy, for example, writes, "Machines as simple as thermostats can be said to have beliefs, and having beliefs seems to be a characteristic of most machines capable of problem solving performance" (McCarthy, 1979). Anyone who thinks strong AI has a chance as a theory of the mind ought to ponder the implications of that remark. We are asked to accept it as a discovery of strong AI that the hunk of metal on the wall that we use to regulate the temperature has beliefs in exactly the same sense that we, our spouses, and our children have beliefs, and furthermore that "most" of the other machines in the room—telephone, tape recorder, adding machine, electric light switch—also have beliefs in this literal sense. It is not the aim of this article to argue against McCarthy's point, so I will simply assert the following without argument. The study of the mind starts with such facts as that humans have beliefs, while thermostats, telephones, and adding machines don't. If you get a theory that denies this point you have produced a counter example to the theory and the theory is false. One gets the impression that people in AI who write this

sort of thing think they can get away with it because they don't really take it seriously, and they don't think anyone else will either. I propose, for a moment at least, to take it seriously. Think hard for one minute about what would be necessary to establish that that hunk of metal on the wall over there had real beliefs, beliefs with direction of fit, propositional content, and conditions of satisfaction; beliefs that had the possibility of being strong beliefs or weak beliefs; nervous, anxious, or secure beliefs; dogmatic, rational, or superstitious beliefs; blind faiths or hesitant cogitations; any kind of beliefs. The thermostat is not a candidate. Neither is stomach, liver, adding machine, or telephone. However, since we are taking the idea seriously, notice that its truth would be fatal to strong AI's claim to be a science of the mind. For now the mind is everywhere. What we wanted to know is what distinguishes the mind from thermostats and livers. And if McCarthy were right, strong AI wouldn't have a hope of telling us that.

5.2 The Robot Reply (Yale)

"Suppose we wrote a different kind of program from Schank's program. Suppose we put a computer inside a robot, and this computer would not just take in formal symbols as input and give out formal symbols as output, but rather would actually operate the robot in such a way that the robot does something very much like perceiving, walking, moving about, hammering nails, eating, drinking—anything you like. The robot would, for example, have a television camera attached to it that enabled it to 'see,' it would have arms and legs that enabled it to 'act,' and all of this would be controlled by its computer 'brain.' Such a robot would, unlike Schank's computer, have genuine understanding and other mental states."

The first thing to notice about the robot reply is that it tacitly concedes that cognition is not solely a matter of formal symbol manipulation, since this reply adds a set of causal relation with the outside world (cf. Fodor, 1980). But the answer to the robot reply is that the addition of such "perceptual" and "motor" capacities adds nothing by way of understanding, in particular, or intentionality, in general, to Schank's original program. To see this, notice that the same thought experiment applies to the robot case. Suppose that instead of the computer inside the robot, you put me inside the room and, as in the original Chinese case, you give me more Chinese symbols with more instructions in English for matching Chinese symbols to Chinese symbols and feeding back Chinese symbols to the outside. Suppose, unknown to me, some of the Chinese symbols that come to me come from a television camera attached to the robot and other Chinese symbols that I am giving out serve to make the motors inside the robot move the robot's legs or arms. It is important to emphasize that all I am doing is manipulating formal symbols: I know none of these other facts. I am receiving "information" from the robot's "perceptual" apparatus, and I am giving out "instructions" to its motor apparatus without knowing either of these facts. I am the robot's homunculus, but unlike the traditional homunculus, I don't know what's going on. I don't understand anything except the rules for symbol manipulation. Now in this case I want to say that the robot has no intentional

states at all; it is simply moving about as a result of its electrical wiring and its program. And furthermore, by instantiating the program I have no intentional states of the relevant type. All I do is follow formal instructions about manipulating formal-symbols.

5.3 The Brain Simulator Reply (Berkeley and M.I.T.)

"Suppose we design a program that doesn't represent information that we have about the world, such as the information in Schank's scripts, but simulates the actual sequence of neuron firings at the synapses of the brain of a native Chinese speaker when he understands stories in Chinese and gives answers to them. The machine takes in Chinese stories and questions about them as input, it simulates the formal structure of actual Chinese brains in processing these stories, and it gives out Chinese answers as outputs. We can even imagine that the machine operates, not with a single serial program, but with a whole set of programs operating in parallel, in the manner that actual human brains presumably operate when they process natural language. Now surely in such a case we would have to say that the machine understood the stories; and if we refuse to say that, wouldn't we also have to deny that native Chinese speakers understood the stories? At the level of the synapses, what would or could be different about the program of the computer and the program of the Chinese brain?"

Before countering this reply I want to digress to note that it is an odd reply for any partisan of artificial intelligence (or functionalism, etc.) to make: I thought the whole idea of strong AI is that we don't need to know how the brain works to know how the mind works. The basic hypothesis, or so I had supposed, was that there is a level of mental operations consisting of computational processes over formal elements that constitute the essence of the mental and can be realized in all sorts of different brain processes, in the same way that any computer program can be realized in different computer hardwares: on the assumptions of strong AI, the mind is to the brain as the program is to the hardware, and thus we can understand the mind without doing neurophysiology. If we had to know how the brain worked to do AI, we wouldn't bother with AI. However, even getting this close to the operation of the brain is still not sufficient to produce understanding. To see this, imagine that instead of a monolingual man in a room shuffling symbols we have the man operate an elaborate set of water pipes with valves connecting them. When the man receives the Chinese symbols, he looks up in the program, written in English, which valves he has to turn on and off. Each water connection corresponds to a synapse in the Chinese brain, and the whole system is rigged up so that after doing all the right firings, that is after turning on all the right faucets, the Chinese answers pop out at the output end of the series of pipes.

Now where is the understanding in this system? It takes Chinese as input, it simulates the formal structure of the synapses of the Chinese brain, and it gives Chinese as output. But the man certainly doesn't understand Chinese, and neither do the water pipes, and if we are tempted to adopt what I think is the absurd view that somehow the *conjunction* of man *and* water pipes understands,

remember that in principle the man can internalize the formal structure of the water pipes and do all the "neuron firings" in his imagination. The problem with the brain simulator is that it is simulating the wrong things about the brain. As long as it simulates only the formal structure of the sequence of neuron firings at the synapses, it won't have simulated what matters about the brain, namely its causal properties, its ability to produce intentional states. And that the formal properties are not sufficient for the causal properties is shown by the water pipe example: we can have all the formal properties carved off from the relevant neurobiological causal properties.

5.4 The Combination Reply (Berkeley and Stanford)

"While each of the previous three replies might not be completely convincing by itself as a refutation of the Chinese room counterexample, if you take all three together they are collectively much more convincing and even decisive. Imagine a robot with a brain-shaped computer lodged in its cranial cavity, imagine the computer programmed with all the synapses of a human brain, imagine the whole behavior of the robot is indistinguishable from human behavior, and now think of the whole thing as a unified system and not just as a computer with inputs and outputs. Surely in such a case we would have to ascribe intentionality to the system."

I entirely agree that in such a case we would find it rational and indeed irresistible to accept the hypothesis that the robot had intentionality, as long as we knew nothing more about it. Indeed, besides appearance and behavior, the other elements of the combination are really irrelevant. If we could build a robot whose behavior was indistinguishable over a large range from human behavior, we would attribute intentionality to it, pending some reason not to. We wouldn't need to know in advance that its computer brain was a formal analogue of the human brain.

But I really don't see that this is any help to the claims of strong AI; and here's why: According to strong AI, instantiating a formal program with the right input and output is a sufficient condition of, indeed is constitutive of, intentionality. As Newell (1979) puts it, the essence of the mental is the operation of a physical symbol system. But the attributions of intentionality that we make to the robot in this example have nothing to do with formal programs. They are simply based on the assumption that if the robot looks and behaves sufficiently like us, then we would suppose, until proven otherwise, that it must have mental states like ours that cause and are expressed by its behavior and it must have an inner mechanism capable of producing such mental states. If we knew independently how to account for its behavior without such assumptions we would not attribute intentionality to it, especially if we knew it had a formal program. And this is precisely the point of my earlier reply to the objection in section 5.2.

Suppose we knew that the robot's behavior was entirely accounted for by the fact that a man inside it was receiving uninterpreted formal symbols from the robot's sensory receptors and sending out uninterpreted formal symbols to its motor mechanisms, and the man was doing this symbol manipulation in

accordance with a bunch of rules. Furthermore, suppose the man knows none of these facts about the robot, all he knows is which operations to perform on which meaningless symbols. In such a case we would regard the robot as an ingenious mechanical dummy. The hypothesis that the dummy has a mind would now be unwarranted and unnecessary, for there is now no longer any reason to ascribe intentionality to the robot or to the system of which it is a part (except of course for the man's intentionality in manipulating the symbols). The formal symbol manipulations go on, the input and output are correctly matched, but the only real locus of intentionality is the man, and he doesn't know any of the relevant intentional states; he doesn't, for example, *see* what comes into the robot's eyes, he doesn't *intend* to move the robot's arm, and he doesn't *understand* any of the remarks made to or by the robot. Nor, for the reasons stated earlier, does the system of which man and robot are a part.

To see this point, contrast this case with cases in which we find it completely natural to ascribe intentionality to members of certain other primate species such as apes and monkeys and to domestic animals such as dogs. The reasons we find it natural are, roughly, two: we can't make sense of the animal's behavior without the ascription of intentionality, and we can see that the beasts are made of similar stuff to ourselves—that is an eye, that a nose, this is its skin, and so on. Given the coherence of the animal's behavior and the assumption of the same causal stuff underlying it, we assume both that the animal must have mental states underlying its behavior, and that the mental states must be produced by mechanisms made out of the stuff that is like our stuff. We would certainly make similar assumptions about the robot unless we had some reason not to, but as soon as we knew that the behavior was the result of a formal program, and that the actual causal properties of the physical substance were irrelevant we would abandon the assumption of intentionality (See "Cognition and Consciousness in Nonhuman Species," *The Behavioral and Brain Sciences* (1978), 1 (4)).

There are two other responses to my example that come up frequently (and so are worth discussing) but really miss the point.

5.5 The Other Minds Reply (Yale)

"How do you know that other people understand Chinese or anything else? Only by their behavior. Now the computer can pass the behavioral tests as well as they can (in principle), so if you are going to attribute cognition to other people you must in principle also attribute it to computers."

This objection really is only worth a short reply. The problem in this discussion is not about how I know that other people have cognitive states, but rather what it is that I am attributing to them when I attribute cognitive states to them. The thrust of the argument is that it couldn't be just computational processes and their output because the computational processes and their output can exist without the cognitive state. It is no answer to this argument to feign anesthesia. In "cognitive sciences" one presupposes the reality and knowability of the mental in the same way that in physical sciences one has to presuppose the reality and knowability of physical objects.

5.6 The Many Mansions Reply (Berkeley)

"Your whole argument presupposes that AI is only about analogue and digital computers. But that just happens to be the present state of technology. Whatever these causal processes are that you say are essential for intentionality (assuming you are right), eventually we will be able to build devices that have these causal processes, and that will be artificial intelligence. So your arguments are in no way directed at the ability of artificial intelligence to produce and explain cognition."

I really have no objection to this reply save to say that it in effect trivializes the project of strong AI by redefining it as whatever artificially produces and explains cognition. The interest of the original claim made on behalf of artificial intelligence is that it was a precise, well defined thesis: mental processes are computational processes over formally defined elements. I have been concerned to challenge that thesis. If the claim is redefined so that it is no longer that thesis, my objections no longer apply because there is no longer a testable hypothesis for them to apply to.

Let us now return to the question I promised I would try to answer: granted that in my original example I understand the English and I do not understand the Chinese, and granted therefore that the machine doesn't understand either English or Chinese, still there must be something about me that makes it the case that I understand English and a corresponding something lacking in me that makes it the case that I fail to understand Chinese. Now why couldn't we give those somethings, whatever they are, to a machine?

I see no reason in principle why we couldn't give a machine the capacity to understand English or Chinese, since in an important sense our bodies with our brains are precisely such machines. But I do see very strong arguments for saying that we could not give such a thing to a machine where the operation of the machine is defined solely in terms of computational processes over formally defined elements; that is, where the operation of the machine is defined as an instantiation of a computer program. It is not because I am the instantiation of a computer program that I am able to understand English and have other forms of intentionality (I am, I suppose, the instantiation of any number of computer programs), but as far as we know it is because I am a certain sort of organism with a certain biological (i.e. chemical and physical) structure, and this structure, under certain conditions, is causally capable of producing perception, action, understanding, learning, and other intentional phenomena. And part of the point of the present argument is that only something that had those causal powers could have that intentionality. Perhaps other physical and chemical processes could produce exactly these effects; perhaps, for example, Martians also have intentionality but their brains are made of different stuff. That is an empirical question, rather like the question whether photosynthesis can be done by something with a chemistry different from that of chlorophyll.

But the main point of the present argument is that no purely formal model will ever be sufficient by itself for intentionality because the formal properties are not by themselves constitutive of intentionality, and they have by themselves no causal powers except the power, when instantiated, to produce the next stage of the formalism when the machine is running. And any other causal

properties that particular realizations of the formal model have, are irrelevant to the formal model because we can always put the same formal model in a different realization where those causal properties are obviously absent. Even if, by some miracle, Chinese speakers exactly realize Schank's program, we can put the same program in English speakers, water pipes, or computers, none of which understand Chinese, the program notwithstanding.

What matters about brain operations is not the formal shadow cast by the sequence of synapses but rather the actual properties of the sequences. All the arguments for the strong version of artificial intelligence that I have seen insist on drawing an outline around the shadows cast by cognition and then claiming that the shadows are the real thing.

By way of concluding I want to try to state some of the general philosophical points implicit in the argument. For clarity I will try to do it in a question and answer fashion, and I begin with that old chestnut of a question:

"Could a machine think?"

The answer is, obviously, yes. We are precisely such machines.

"Yes, but could an artifact, a man-made machine, think?"

Assuming it is possible to produce artificially a machine with a nervous system, neurons with axons and dendrites, and all the rest of it, sufficiently like ours, again the answer to the question seems to be obviously, yes. If you can exactly duplicate the causes, you could duplicate the effects. And indeed it might be possible to produce consciousness, intentionality, and all the rest of it using some other sorts of chemical principles than those that human beings use. It is, as I said, an empirical question.

"OK, but could a digital computer think?"

If by "digital computer" we mean anything at all that has a level of description where it can correctly be described as the instantiation of a computer program, then again the answer is, of course, yes, since we are the instantiations of any number of computer programs, and we can think.

"But could something think, understand, and so on *solely* in virtue of being a computer with the right sort of program? Could instantiating a program, the right program of course, by itself be a sufficient condition of understanding?"

This I think is the right question to ask, though it is usually confused with one or more of the earlier questions, and the answer to it is no.

"Why not?"

Because the formal symbol manipulations by themselves don't have any intentionality; they are quite meaningless; they aren't even *symbol* manipulations, since the symbols don't symbolize anything. In the linguistic jargon, they have only a syntax but no semantics. Such intentionality as computers appear to have is solely in the minds of those who program them and those who use them, those who send in the input and those who interpret the output.

The aim of the Chinese room example was to try to show this by showing that as soon as we put something into the system that really does have intentionality (a man), and we program him with the formal program, you can see that the formal program carries no additional intentionality. It adds nothing, for example, to a man's ability to understand Chinese.

Precisely that feature of AI that seemed so appealing—the distinction between the program and the realization—proves fatal to the claim that simulation could

be duplication. The distinction between the program and its realization in the hardware seems to be parallel to the distinction between the level of mental operations and the level of brain operations. And if we could describe the level of mental operations as a formal program, then it seems we could describe what was essential about the mind without doing either introspective psychology or neurophysiology of the brain. But the equation, "mind is to brain as program is to hardware" breaks down at several points, among them the following three:

First, the distinction between program and realization has the consequence that the same program could have all sorts of crazy realizations that had no form of intentionality. Weizenbaum (1976; ch. 2), for example, shows in detail how to construct a computer using a roll of toilet paper and a pile of small stones. Similarly, the Chinese story understanding program can be programmed into a sequence of water pipes, a set of wind machines, or a monolingual English speaker, none of which thereby acquires an understanding of Chinese. Stones, toilet paper, wind, and water pipes are the wrong kind of stuff to have intentionality in the first place—only something that has the same causal powers as brains can have intentionality—and though the English speaker has the right kind of stuff for intentionality you can easily see that he doesn't get any extra intentionality by memorizing the program, since memorizing it won't teach him Chinese.

Second, the program is purely formal, but the intentional states are not in that way formal. They are defined in terms of their content, not their form. The belief that it is raining, for example, is not defined as a certain formal shape, but as a certain mental content with conditions of satisfaction, a direction of fit (see Searle, 1979b), and the like. Indeed the belief as such hasn't even got a formal shape in this syntactic sense, since one and the same belief can be given an indefinite number of different syntactic expressions in different linguistic systems.

Third, as I mentioned before, mental states and events are literally a product of the operation of the brain, but the program is not in that way a product of the computer.

"Well if programs are in no way constitutive of mental processes, why have so many people believed the converse? That at least needs some explanation."

I don't really know the answer to that one. The idea that computer simulations could be the real thing ought to have seemed suspicious in the first place because the computer isn't confined to simulating mental operations, by any means. No one supposes that computer simulations of a five-alarm fire will burn the neighborhood down or that a computer simulation of a rainstorm will leave us all drenched. Why on earth would anyone suppose that a computer simulation of understanding actually understood anything? It is sometimes said that it would be frightfully hard to get computers to feel pain or fall in love, but love and pain are neither harder nor easier than cognition or anything else. For simulation, all you need is the right input and output and a program in the middle that transforms the former into the latter. That is all the computer has for anything it does. To confuse simulation with duplication is the same mistake, whether it is pain, love, cognition, fires, or rainstorms.

Still, there are several reasons why AI must have seemed—and to many people perhaps still does seem—in some way to reproduce and thereby explain mental phenomena, and I believe we will not succeed in removing these illusions until we have fully exposed the reasons that give rise to them.

First, and perhaps most important, is a confusion about the notion of "information processing": many people in cognitive science believe that the human brain, with its mind, does something called "information processing," and analogously the computer with its program does information processing; but fires and rainstorms, on the other hand, don't do information processing at all. Thus, though the computer can simulate the formal features of any process whatever, it stands in a special relation to the mind and brain because when the computer is properly programmed, ideally with the same program as the brain, the information processing is identical in the two cases, and this information processing is really the essence of the mental. But the trouble with this argument is that it rests on an ambiguity in the notion of "information." In the sense in which people "process information" when they reflect, say, on problems in arithmetic or when they read and answer questions about stories, the programmed computer does not do "information processing." Rather, what it does is manipulate formal symbols. The fact that the programmer and the interpreter of the computer output use the symbols to stand for objects in the world is totally beyond the scope of the computer. The computer, to repeat, has a syntax but no semantics. Thus, if you type into the computer "2 plus 2 equals?" it will type out "4." But it has no idea that "4" means 4 or that it means anything at all. And the point is not that it lacks some second-order information about the interpretation of its first-order symbols, but rather that its first-order symbols don't have any interpretations as far as the computer is concerned. All the computer has is more symbols. The introduction of the notion of "information processing" therefore produces a dilemma: either we construe the notion of "information processing" in such a way that it implies intentionality as part of the process or we don't. If the former, then the programmed computer does not do information processing, it only manipulates formal symbols. If the latter, then, though the computer does information processing, it is only doing so in the sense in which adding machines, typewriters, stomachs, thermostats, rainstorms, and hurricanes do information processing; namely, they have a level of description at which we can describe them as taking information in at one end, transforming it, and producing information as output. But in this case it is up to outside observers to interpret the input and output as information in the ordinary sense. And no similarity is established between the computer and the brain in terms of any similarity of information processing.

Second, in much of AI there is a residual behaviorism or operationalism. Since appropriately programmed computers can have input–output patterns similar to those of human beings, we are tempted to postulate mental states in the computer similar to human mental states. But once we see that it is both conceptually and empirically possible for a system to have human capacities in some realm without having any intentionality at all, we should be able to overcome this impulse. My desk adding machine has calculating capacities, but no intentionality, and in this paper I have tried to show that a system could

have input and output capabilities that duplicated those of a native Chinese speaker and still not understand Chinese, regardless of how it was programmed. The Turing test is typical of the tradition in being unashamedly behavioristic and operationalistic, and I believe that if AI workers totally repudiated behaviorism and operationalism much of the confusion between simulation and duplication would be eliminated.

Third, this residual operationalism is joined to a residual form of dualism; indeed strong AI only makes sense given the dualistic assumption that, where the mind is concerned, the brain doesn't matter. In strong AI (and in functionalism, as well) what matters are programs, and programs are independent of their realization in machines; indeed, as far as AI is concerned, the same program could be realized by an electronic machine, a Cartesian mental substance, or a Hegelian world spirit. The single most surprising discovery that I have made in discussing these issues is that many AI workers are quite shocked by my idea that actual human mental phenomena might be dependent on actual physical-chemical properties of actual human brains. But if you think about it a minute you can see that I should not have been surprised; for unless you accept some form of dualism, the strong AI project hasn't got a chance. The project is to reproduce and explain the mental by designing programs, but unless the mind is not only conceptually but empirically independent of the brain you couldn't carry out the project, for the program is completely independent of any realization. Unless you believe that the mind is separable from the brain both conceptually and empirically—dualism in a strong form—you cannot hope to reproduce the mental by writing and running programs since programs must be independent of brains or any other particular forms of instantiation. If mental operations consist in computational operations on formal symbols, then it follows that they have no interesting connection with the brain; the only connection would be that the brain just happens to be one of the indefinitely many types of machines capable of instantiating the program. This form of dualism is not the traditional Cartesian variety that claims there are two sorts of *substances*, but it is Cartesian in the sense that it insists that what is specifically mental about the mind has no intrinsic connection with the actual properties of the brain. This underlying dualism is masked from us by the fact that AI literature contains frequent fulminations against "dualism"; what the authors seem to be unaware of is that their position presupposes a strong version of dualism.

"Could a machine think?" My own view is that *only* a machine could think, and indeed only very special kinds of machines, namely brains and machines that had the same causal powers as brains. And that is the main reason strong AI has had little to tell us about thinking, since it has nothing to tell us about machines. By its own definition, it is about programs, and programs are not machines. Whatever else intentionality is, it is a biological phenomenon, and it is as likely to be as causally dependent on the specific biochemistry of its origins as lactation, photosynthesis, or any other biological phenomena. No one would suppose that we could produce milk and sugar by running a computer simulation of the formal sequences in lactation and photosynthesis, but where the mind is concerned many people are willing to believe in such a miracle because of a deep and abiding dualism: the mind they suppose is a matter of for-

mal processes and is independent of quite specific material causes in the way that milk and sugar are not.

In defense of this dualism the hope is often expressed that the brain is a digital computer (early computers, by the way, were often called "electronic brains"). But that is no help. Of course the brain is a digital computer. Since everything is a digital computer, brains are too. The point is that the brain's causal capacity to produce intentionality cannot consist in its instantiating a computer program, since for any program you like it is possible for something to instantiate that program and still not have any mental states. Whatever it is that the brain does to produce intentionality, it cannot consist in instantiating a program since no program, by itself, is sufficient for intentionality.

Acknowledgments

I am indebted to a rather large number of people for discussion of these matters and for their patient attempts to overcome my ignorance of artificial intelligence. I would especially like to thank Ned Block, Hubert Dreyfus, John Haugeland, Roger Schank, Robert Wilensky, and Terry Winograd.

Notes

1. I am not, of course, saying that Schank himself is committed to these claims.
2. Also, "understanding" implies both the possession of mental (intentional) states and the truth (validity, success) of these states. For the purposes of this discussion we are concerned only with the possession of the states.
3. Intentionality is by definition that feature of certain mental states by which they are directed at or about objects and states of affairs in the world. Thus, beliefs, desires, and intentions are intentional states; undirected forms of anxiety and depression are not. For further discussion see Searle (1979b).

References

Fodor, J. A. (1980) Methodological solipsism considered as a research strategy in cognitive psychology. *The Behavioral and Brain Sciences*, 3: 1.

McCarthy, J. (1979) Ascribing mental qualities to machines. In *Philosophical Perspectives in Artificial Intelligence*, ed. M. Ringle. Atlantic Highlands, NJ: Humanities Press.

Newell, A. (1979) Physical symbol systems. Lecture at the La Jolla Conference on Cognitive Science.

Newell, A., and Simon, H. A. (1963) GPS, a program that simulates human thought. In *Computers and Thought*, ed. A. Feigenbaum and V. Feldman, pp. 279–93. New York: McGraw-Hill.

Pylyshyn, Z. W. (1980) Computation and cognition: issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, 3: 1.

Schank, R. C., and Abelson, R. P. (1977) *Scripts, Plans, Coals, and Understanding*. Hillsdale, NJ: Lawrence Erlbaum.

Searle, J. R. (1979a) Intentionality and the use of language. In *Meaning and Use*, ed. A. Margalit. Dordrecht: Reidel.

Searle, J. R. (1979b) What is an intentional state? *Mind*, 88, 74–92.

Weizenbaum, J. (1965) ELIZA—a computer program for the study of natural language communication between man and machine. *Communication of the Association for Computing Machinery* 9, 36–45.

Weizenbaum, J. (1976) *Computer Power and Human Reason*. San Francisco: W. H. Freeman.

Winograd, T. (1973) A procedural model of language understanding. In *Computer Models of Thought and Language*, ed. R. Schank and K. Colby. San Francisco: W. H. Freeman.

PART IV

Experimental Design

Chapter 6

Experimental Design in Psychological Research

Daniel J. Levitin

6.1 Introduction

Experimental design is a vast topic. As one thinks about the information derived from scientific studies, one confronts difficult issues in statistical theory and the limits of knowledge. In this chapter, we confine our discussion to a few of the most important issues in experimental design. This will enable students with no background in behavior research to critically evaluate psychological experiments, and to better understand the nature of empirical research in cognitive science.

Experimental psychology is a young science. The first laboratory of experimental psychology was established just over 100 years ago. Consequently, there are a great many mysteries about human behavior, perception, and performance that have not yet been solved. This makes it an exciting time to engage in psychological research—the field is young enough that there is still a great deal to do, and it is not difficult to think up interesting experiments. The goal of this chapter is to guide the reader in planning and implementing experiments, and in thinking about good experimental design.

A "good" experiment is one in which variables are carefully controlled or accounted for so that one can draw reasonable conclusions from the experiment's outcome.

6.2 The Goals of Scientific Research

Generally, scientific research has four goals:

1. Description of behavior
2. Prediction of behavior
3. Determination of the causes of behavior
4. Explanations of behavior

These goals apply to the physical sciences as well as to the behavioral and life sciences. In basic science, the researcher's primary concern is not with applications for a given finding. The goal of basic research is to increase our understanding of how the world works, or how things came to be the way they are.

Describing behavior impartially is the foremost task of the descriptive study, and because this is never completely possible, one tries to document any

From "Experimental Design in Psychoacoustic Research," chapter 23 in *Music, Cognition, and Computerized Sound* (Cambridge, MA: MIT Press, 1999), 299–328. Reprinted with permission.

systematic biases that could influence descriptions (goal 1). By studying a phenomenon, one frequently develops the ability to *predict* certain behaviors or outcomes (goal 2), although prediction is possible without an understanding of underlying causes (we'll look at some examples in a moment). Controlled experiments are one tool that scientists use to reveal underlying causes so that they can advance from merely predicting behavior to understanding the *cause* of behavior (goal 3). *Explaining* behavior (goal 4) requires more than just a knowledge of causes; it requires a detailed understanding of the mechanisms by which the causal factors perform their functions.

To illustrate the distinction between the four goals of scientific research, consider the history of astronomy. The earliest astronomers were able to *describe* the positions and motions of the stars in the heavens, although they had no ability to *predict* where a given body would appear in the sky at a future date. Through careful observations and documentation, later astronomers became quite skillful at *predicting* planetary and stellar motion, although they lacked an understanding of the underlying factors that *caused* this motion. Newton's laws of motion and Einstein's special and general theories of relativity, taken together, showed that gravity and the contour of the space–time continuum cause the motions we observe. Precisely how gravity and the topology of space–time accomplish this still remains unclear. Thus, astronomy has advanced to the determination of causes of stellar motion (goal 3), although a full *explanation* remains elusive. That is, saying that gravity is responsible for astronomical motion only puts a name on things; it does not tell us how gravity actually works.

As an illustration from behavioral science, one might note that people who listen to loud music tend to lose their high-frequency hearing (description). Based on a number of observations, one can predict that individuals with normal hearing who listen to enough loud music will suffer hearing loss (prediction). A controlled experiment can determine that the loud music is the cause of the hearing loss (determining causality). Finally, study of the cochlea and basilar membrane, and observation of damage to the delicate hair cells after exposure to high-pressure sound waves, meets the fourth goal (explanation).

6.3 Three Types of Scientific Studies

In science there are three broad classes of studies: controlled studies, correlational studies, and descriptive studies. Often the type of study you will be able to do is determined by practicality, cost, or ethics, not directly by your own choice.

6.3.1 Controlled Studies ("True Experiments")

In a controlled experiment, the researcher starts with a group of subjects and randomly assigns them to an experimental condition. The point of *random assignment* is to control for extraneous variables that might affect the outcome of the experiment: variables that are different from the variable(s) being studied. With random assignment, one can be reasonably certain that any differences among the experimental groups were caused by the variable(s) manipulated in the experiment.

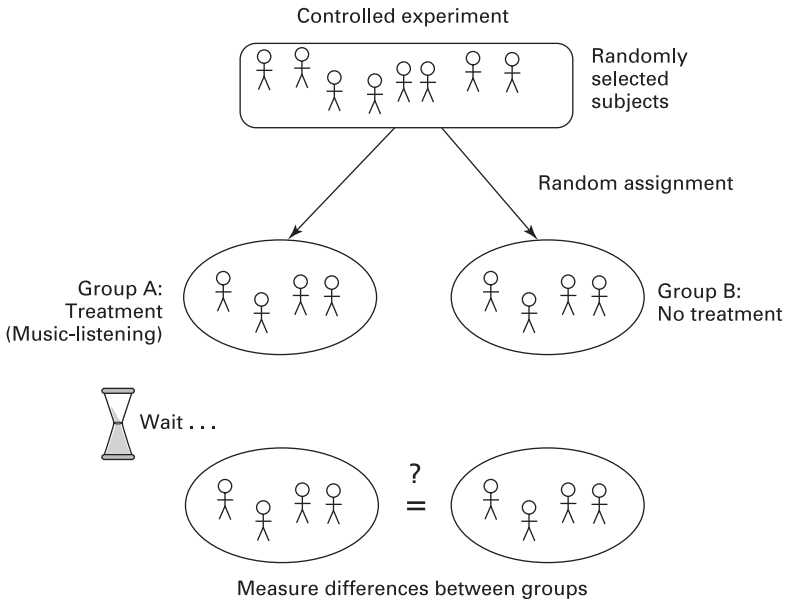


Figure 6.1
In a controlled experiment, subjects are randomly assigned to conditions, and differences between groups are measured.

A controlled experiment in medical research might seek to discover if a certain food additive causes cancer. The researcher might randomly divide a group of laboratory mice into two smaller groups, giving the food additive to one group and not to the other. The variable he/she is interested in is the effect of the food additive; in the language of experimental design, this is called the "independent variable." After a period of time, the researcher compares the mortality rates of the two groups; this quantity is called the "dependent variable" (figure 6.1). Suppose the group that received the additive tended to die earlier. In order to deduce that the additive caused the difference between the groups, the conditions must have been identical in every other respect. Both groups should have had the same diet, same feeding schedule, same temperature in their cages, and so on. Furthermore, the two groups of mice should have started out with similar characteristics, such as age, sex, and so on, so that these variables—being equally distributed between the two groups—can be ruled out as possible causes of the difference in mortality rates.

The two key components of a controlled experiment are *random assignment* of subjects, and *identical experimental conditions* (see figure 6.1). A researcher might have a hypothesis that people who study for an exam while listening to music will score better than people who study in silence. In the language of experimental design, music-listening is the *independent variable*, and test performance, the quantity to be measured, is the *dependent variable*.

No one would take this study seriously if the subjects were divided into two groups based on how they did on the previous exam—if, for instance, the top half of the students were placed in the music-listening condition, and the

bottom half of the students in the silence condition. Then if the result of the experiment was that the music listeners as a group tended to perform better on their next exam, one could argue that this was not because they listened to music while they studied, but because they were the better students to begin with.

Again, the theory behind random assignment is to have groups of subjects who start out the same. Ideally, each group will have similar distributions on every conceivable dimension—age, sex, ethnicity, IQ, and variables that you might not think are important, such as handedness, astrological sign, or favorite television show. Random assignment makes it unlikely that there will be any large systematic differences between the groups.

A similar design flaw would arise if the *experimental conditions* were different. For example, if the music-listening group studied in a well-lit room with windows, and the silence group studied in a dark, windowless basement, any difference between the groups could be due to the different environments. The room conditions become confounded with the music-listening conditions, such that it is impossible to deduce which of the two is the causal factor.

Performing random assignment of subjects is straightforward. Conceptually, one wants to mix the subjects' names or numbers thoroughly, then draw them out of a hat. Realistically, one of the easiest ways to do this is to generate a different random number for each subject, and then sort the random numbers. If n equals the total number of subjects you have, and g equals the number of groups you are dividing them into, the first n/g subjects will comprise the first group, the next n/g will comprise the second group, and so on.

If the results of a controlled experiment indicate a difference between groups, the next question is whether these findings are generalizable. If your initial group of subjects (the large group, before you randomly assigned subjects to conditions) was also randomly selected (called *random sampling* or *random selection*, as opposed to *random assignment*), this is a reasonable conclusion to draw. However, there are almost always some constraints on one's initial choice of subjects, and this constrains generalizability. For example, if all the subjects you studied in your music-listening experiment lived in fraternities, the finding might not generalize to people who do not live in fraternities. If you want to be able to generalize to all college students, you would need to take a representative sample of all college students. One way to do this is to choose your subjects randomly, such that each member of the population you are considering (college students) has an equal likelihood of being placed in the experiment.

There are some interesting issues in representative sampling that are beyond the scope of this chapter. For example, if you wanted to take a representative sample of all American college students and you chose American college students randomly, it is possible that you would be choosing several students from some of the larger colleges, such as the University of Michigan, and you might not choose any students at all from some of the smaller colleges, such as Bennington College; this would limit the applicability of your findings to the colleges that were represented in your sample. One solution is to conduct a *stratified sample*, in which you first randomly select colleges (making it just as likely that you'll choose large and small colleges) and then randomly select the

same number of students from each of those colleges. This ensures that colleges of different sizes are represented in the sample. You then weight the data from each college in accordance with the percentage contribution each college makes to the total student population of your sample. (For further reading, see Shaughnessy and Zechmeister 1994.)

Choosing subjects randomly requires careful planning. If you try to take a random sample of Stanford students by standing in front of the Braun Music Building and stopping every third person coming out, you might be selecting a greater percentage of music students than actually exists on campus. Yet truly random samples are not always practical. Much psychological research is conducted on college students who are taking an introductory psychology class, and are required to participate in an experiment for course credit. It is not at all clear whether American college students taking introductory psychology are representative of students in general, or of people in the world in general, so one should be careful not to overgeneralize findings from these studies.

6.3.2 Correlational Studies

A second type of study is the *correlational study* (figure 6.2). Because it is not always practical or ethical to perform random assignments, scientists are sometimes forced to rely on patterns of co-occurrence, or correlations between events. The classic example of a correlational study is the link between cigarette smoking and cancer. Few educated people today doubt that smokers are more likely to die of lung cancer than are nonsmokers. However, in the history of scientific research there has never been a controlled experiment with human subjects on this topic. Such an experiment would take a group of healthy nonsmokers, and randomly assign them to two groups, a smoking group and a nonsmoking group. Then the experimenter would simply wait until most of the people in the study have died, and compare the average ages and causes of death of the two groups. Because our hypothesis is that smoking causes cancer, it would clearly be unethical to ask people to smoke who otherwise would not.

The scientific evidence we have that smoking causes cancer is correlational. That is, when we look at smokers as a group, a higher percentage of them do indeed develop fatal cancers, and die earlier, than do nonsmokers. But without a controlled study, the possibility exists that there is a third factor—a mysterious "factor x"—that both causes people to smoke and to develop cancer. Perhaps there is some enzyme in the body that gives people a nicotine craving, and this same enzyme causes fatal cancers. This would account for both outcomes, the kinds of people who smoke and the rate of cancers among them, and it would show that there is no causal link between smoking and cancer.

In correlational studies, a great deal of effort is devoted to trying to uncover differences between the two groups studied in order to identify any causal factors that might exist. In the case of smoking, none have been discovered so far, but the failure to discover a third causal factor does not prove that one does not exist. It is an axiom in the philosophy of science that one can prove only the presence of something; one can't prove the absence of something—it could always be just around the corner, waiting to be discovered in the next experiment (Hempel 1966). In the real world, behaviors and diseases are usually brought

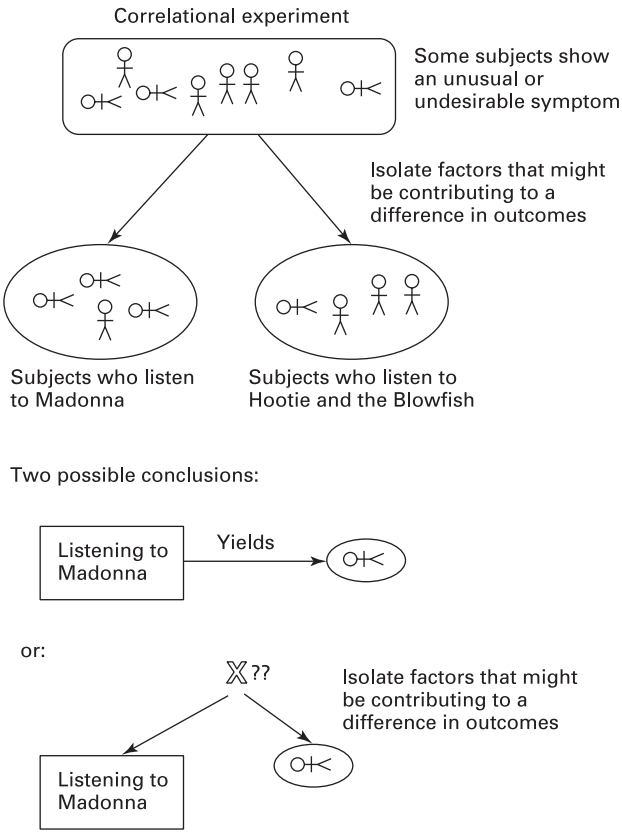


Figure 6.2
In a correlational study, the researcher looks for a relation between two observed behaviors—in this case, the relation between untimely death and listening to Madonna recordings.

on by a number of complicated factors, so the mysterious third variable, "factor x," could in fact be a collection of different, and perhaps unrelated, variables that act together to cause the outcomes we observe.

An example of a correlational study with a hypothesized musical cause is depicted in figure 6.2. Such a study would require extensive interviews with the subjects (or their survivors), to try to determine all factors that might separate the subjects exhibiting the symptom from the subjects without the symptom.

The problem with correlational studies is that the search for underlying factors that account for the differences between groups can be very difficult. Yet many times, correlational studies are all we have, because ethical considerations preclude the use of controlled experiments.

6.3.3 Descriptive Studies

Descriptive studies do not look for differences between people or groups, but seek only to describe an aspect of the world as it is. A descriptive study in physics might seek to discover what elements make up the core of the planet Jupiter. The goal in such a study would not be to compare Jupiter's core with

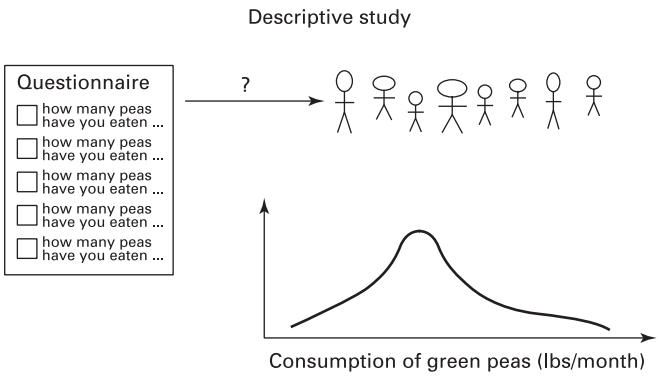


Figure 6.3
In a descriptive study, the researcher seeks to describe some aspect of the state of the world, such as people's consumption of green peas.

the core of other planets, but to learn more about the origins of the universe. In psychology, we might want to know the part of the brain that is activated when someone performs a mental calculation, or the number of pounds of fresh green peas the average Canadian eats in a year (figure 6.3). Our goal in these cases is not to contrast individuals but to acquire some basic data about the nature of things. Of course, descriptive studies can be used to establish "norms," so that we can compare people against the average, but as their name implies, the primary goal in descriptive experiments is often just to describe something that had not been described before. Descriptive studies are every bit as useful as controlled experiments and correlational studies—sometimes, in fact, they are even more valuable because they lay the foundation for further experimental work.

6.4 Design Flaws in Experimental Design

6.4.1 Clever Hans

There are many examples of flawed studies or flawed conclusions that illustrate the difficulties in controlling extraneous variables. Perhaps the most famous case is that of Clever Hans.

Clever Hans was a horse owned by a German mathematics teacher around the turn of the twentieth century. Hans became famous following many demonstrations in which he could perform simple addition and subtraction, read German, and answer simple questions by tapping his hoof on the ground (Watson 1967). One of the first things that skeptics wondered (as you might) is whether Hans would continue to be clever when someone other than his owner asked the questions, or when Hans was asked questions that he had never heard before. In both these cases, Hans continued to perform brilliantly, tapping out the sums or differences for arithmetic problems.

In 1904, a scientific commission was formed to investigate Hans's abilities more carefully. The commission discovered, after rigorous testing, that Hans could never answer a question if the questioner did not also know the answer,

or if Hans could not see his questioner. It was finally discovered that Hans had become very adept at picking up subtle (and probably unintentional) movements on the part of the questioner that cued him as to when he should stop tapping his foot. Suppose a questioner asked Hans to add 7 and 3. Hans would start tapping his hoof, and keep on tapping until the questioner stopped him by saying "Right! Ten!" or, more subtly, by moving slightly when the correct answer was reached.

You can see how important it is to ensure that extraneous cues or biases do not intrude into an experimental situation.

6.4.2 Infants' Perception of Musical Structure

In studies of infants' perception of music, infants typically sit in their mother's lap while music phrases are played over a speaker. Infants tend to turn their heads toward a novel or surprising event, and this is the dependent variable in many infant studies; the point at which the infants turn their heads indicates when they perceive a difference in whatever is being played. Suppose you ran such a study and found that the infants were able to distinguish Mozart selections that were played normally from selections of equal length that began or ended in the middle of a musical phrase. You might take this as evidence that the infants have an innate understanding of musical phraseology.

Are there alternative explanations for the results? Suppose that in the experimental design, the mothers could hear the music, too. The mothers might unconsciously cue the infants to changes in the stimulus that they (the mothers) detect. A simple solution is to have the mothers wear headphones playing white noise, so that their perception of the music is masked.

6.4.3 Computers, Timing, and Other Pitfalls

It is very important that you not take anything for granted as you design a careful experiment, and control extraneous variables. For example, psychologists studying visual perception frequently present their stimuli on a computer using the MacIntosh or Windows operating system. In a computer program, the code may specify that an image is to remain on the computer monitor for a precise number of milliseconds. Just because you specify this does not make it happen, however. Monitors have a *refresh rate* (60 or 75 Hz is typical), so the "on time" of an image will always be an integer multiple of the refresh cycle (13.33 milliseconds for a 75 Hz refresh rate) no matter what you instruct the computer to do in your code. To make things worse, the MacIntosh and Windows operating systems do not guarantee "refresh cycle accuracy" in their updating, so an instruction to put a new image on the screen may be delayed an unknown amount of time.

It is important, therefore, always to verify, using some external means, that the things you think are happening in your experiment are actually happening. Just because you leave the volume control on your amplifier at the same spot doesn't mean the volume of a sound stimulus you are playing will be the same from day to day. You should measure the output and not take the knob position for granted. Just because a frequency generator is set for 1000 Hz does not mean it is putting out a 1000 Hz signal. It is good science for you to measure the output frequency yourself.

6.5 Number of Subjects

How many subjects are enough? In statistics, the word "population" refers to the total group of people to which the researcher wishes to generalize findings. The population might be female sophomores at Stanford, or all Stanford students, or all college students in the United States, or all people in the United States. If one is able to draw a representative sample of sufficient size from a population, one can make inferences about the whole population based on a relatively small number of cases. This is the basis of presidential polls, for example, in which only 2000 voters are surveyed, and the outcome of an election can be predicted with reasonable accuracy.

The size of the sample required is dependent on the degree of homogeneity or heterogeneity in the total population you are studying. In the extreme, if you are studying a population that is so homogeneous that every individual is identical on the dimensions being studied, a sample size of one will provide all the information you need. At the other extreme, if you are studying a population that is so heterogeneous that each individual differs categorically on the dimension you are studying, you will need to sample the entire population.

As a "rough-and-ready" rule, if you are performing a descriptive perceptual experiment, and the phenomenon you are studying is something that you expect to be invariant across people, you need to use only a few subjects, perhaps five. An example of this type of study would be calculating threshold sensitivities for various sound frequencies, such as was done by Fletcher and Munson (1933).

If you are studying a phenomenon for which you expect to find large individual differences, you might need between 30 and 100 subjects. This depends to some degree on how many different conditions there are in the study. In order to obtain means with a relatively small variance, it is a good idea to have at least five to ten subjects in each experimental condition.

6.6 Types of Experimental Designs

Suppose you are researching the effect of music-listening on studying efficiency, as mentioned at the beginning of this chapter. Let's expand on the simpler design described earlier. You might divide your subjects into five groups: two experimental groups and three control groups. One experimental group would listen to rock music, and the other would listen to classical music. Of the three control groups, one would listen to rock music for the same number of minutes per day as the experimental group listening to rock (but not while they were studying); a second would do the same for classical music; the third would listen to no music at all. This is called a *between-subjects* design, because each subject is in one condition and one condition only (also referred to as an *independent groups* design). If you assign 10 subjects to each experimental condition, this would require a total of 50 subjects. Table 6.1 shows the layout of this experiment. Each distinct box in the table is called a *cell* of the experiment, and subject numbers are filled in for each cell. Notice the asymmetry for the *no music* condition. The experiment was designed so that there is only one "no music" condition, whereas there are four music conditions of various types.

Table 6.1
Between-subjects experiment on music and study habits

Condition	Only while studying	Only while not studying
Music		
Classical	Subjects 1–10	Subjects 11–20
Rock	Subjects 21–30	Subjects 31–40
No music	Subjects 41–50	Subjects 41–50

Testing 50 subjects might not be practical. An alternative is a *within-subjects* design, in which every subject is tested in every condition (also called a *repeated measures* design). In this example, a total of ten subjects could be randomly divided into the five conditions, so that two subjects experience each condition for a given period of time. Then the subjects switch to another condition. By the time the experiment is completed, ten observations have been collected in each cell, and only ten subjects are required.

The advantage of each subject experiencing each condition is that you can obtain measures of how each individual is affected by the manipulation, something you cannot do in the between-subjects design. It might be the case that some people do well in one type of condition and other people do poorly in it, and the within-subjects design is the best way to show this. The obvious advantage to the within-subjects design is the smaller number of subjects required. But there are disadvantages as well.

One disadvantage is *demand characteristics.* Because each subject experiences each condition, they are not as naive about the experimental manipulation. Their performance could be influenced by a conscious or unconscious desire to make one of the conditions work better. Another problem is *carryover effects.* Suppose you were studying the effect of Prozac on learning, and that the half-life of the drug is 48 hours. The group that gets the drug first might still be under its influence when they are switched to the nondrug condition. This is a carryover effect. In the music-listening experiment, it is possible that listening to rock music creates anxiety or exhilaration that might last into the next condition.

A third disadvantage of within-subjects designs is *order effects*, and these are particularly troublesome in psychophysical experiments. An order effect is similar to a carryover effect, and it concerns how responses in an experiment might be influenced by the order in which the stimuli or conditions are presented. For instance, in studies of speech discrimination, subjects can habituate (become used to, or become more sensitive) to certain sounds, altering their threshold for the discriminability of related sounds. A subject who habituates to a certain sound may respond differently to the sound immediately following it than he/she normally would. For these reasons, it is important to counterbalance the order of presentations; presenting the same order to every subject makes it difficult to account for any effects that are due merely to order.

One way to reduce order effects is to present the stimuli or conditions in random order. In some studies, this is sufficient, but to be really careful about order effects, the random order simply is not rigorous enough. The solution is to use every possible order. In a *within-subjects* design, each subject would

complete the experiment with each order. In a *between-subjects* design, different subjects would be assigned different orders. The choice will often depend on the available resources (time and availability of subjects). The number of possible orders is *N!* ("n factorial"), where *N* equals the number of stimuli. With two stimuli there are two possible orders ($2! = 2 \times 1$); with three stimuli there are six possible orders ($3! = 3 \times 2 \times 1$); with six stimuli there are 720 possible orders ($6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1$). Seven hundred twenty orders is not practical for a within-subjects design, or for a between-subjects design. One solution in this case is to create an order that presents each stimulus in each serial position. A method for accomplishing this involves using the Latin Square. For even-numbered N, the size of the Latin Square will be $N \times N$; therefore, with six stimuli you would need only 36 orders, not 720. For odd-numbered N, the size of the Latin Square will be $N \times 2N$. Details of this technique are covered in experimental design texts such as Kirk (1982) and Shaughnessy and Zechmeister (1994).

6.7 Ethical Considerations in Using Human Subjects

Some experiments on human subjects in the 1960s and 1970s raised questions about how human subjects are treated in behavioral experiments. As a result, guidelines for human experimentation were established. The American Psychological Association, a voluntary organization of psychologists, formulated a code of ethical principles (American Psychological Association 1992). In addition, most universities have established committees to review and approve research using human subjects. The purpose of these committees is to ensure that subjects are treated ethically, and that fair and humane procedures are followed. In some universities, experiments performed for course work or experiments done as "pilot studies" do not require approval, but these rules vary from place to place, so it is important to determine the requirements at your institution before engaging in any human subject research.

It is also important to understand the following four basic principles of ethics in human subject research:

1. *Informed consent.* Before agreeing to participate in an experiment, subjects should be given an accurate description of their task in the experiment, and told any risks involved. Subjects should be allowed to decline, or to discontinue participation in the experiment at any time without penalty.
2. *Debriefing.* Following the experiment, the subjects should be given an explanation of the hypothesis being tested and the methods used. The experimenter should answer any questions the subjects have about the procedure or hypothesis. Many psychoacoustic experiments involve difficult tasks, leading some subjects to feel frustrated or embarrassed. Subjects should never leave an experiment feeling slow, stupid, or untalented. It is the experimenter's responsibility to ensure that the subjects understand that these tasks are inherently difficult, and when appropriate, the subjects should be told that the data are not being used to evaluate them personally, but to collect information on how the population in general can perform the task.

3. *Privacy and confidentiality.* The experimenter must carefully guard the data that are collected and, whenever possible, code and store the data in such a way that subjects' identities remain confidential.
4. *Fraud.* This principle is not specific to human subjects research, but applies to all research. An essential ethical standard of the scientific community is that scientific researchers never fabricate data, and never knowingly, intentionally, or through carelessness allow false data, analyses, or conclusions to be published. Fraudulent reporting is one of the most serious ethical breaches in the scientific community.

6.8 Analyzing Your Data

6.8.1 Quantitative Analysis

Measurement Error Whenever you measure a quantity, there are two components that contribute to the number you end up with: the actual value of the thing you are measuring and some amount of measurement error, both human and mechanical. It is an axiom of statistics that measurement error is just as likely to result in an overestimate as an underestimate of the true value. That is, each time you take a measurement, the error term (let's call it *epsilon*) is just as likely to be positive as negative. Over a large number of measurements, the positive errors and negative errors will cancel out, and the average value of epsilon will approach 0. The larger the number of measurements you make, the closer you will get to the true value. Thus, as the number of measurements approaches infinity, the arithmetic average of your measurements approaches the true quantity being measured. Suppose we are measuring the weight of a sandbag.

Formally, we would write:

$$n \to \infty, \quad \bar{\varepsilon} = 0$$

where $\bar{\varepsilon}$ = the mean of epsilon, and

$$n \to \infty, \quad \overline{w} = \mathrm{w}$$

where $\overline{w}$ = the mean of all the weight measurements and w = the true weight.

When measuring the behavior of human subjects on a task, you encounter not only measurement error but also performance error. The subjects will not perform identically every time. As with measurement error, the more observations you make, the more likely it is that the performance errors cancel each other out. In psychoacoustic tasks the performance errors can often be relatively large. This is the reason why one usually wants to have the subject perform the same task many times, or to have many subjects perform the task a few times.

Because of these errors, the value of your dependent variable(s) at the end of the experiment will always deviate from the true value by some amount. Statistical analysis helps in interpreting these differences (Bayesian inferencing, meta-analyses, effect size analysis, significance testing) and in predicting the true value (point estimates and confidence intervals). The mechanics of these

tests are beyond the scope of this chapter, and the reader is referred to the statistics textbooks mentioned earlier.

Significance Testing Suppose you wish to observe differences in interval identification ability between brass players and string players. The question is whether the difference you observe between the two groups can be wholly accounted for by measurement and performance error, or whether a difference of the size you observe indicates a true difference in the abilities of these musicians.

Significance tests provide the user with a "p value," the probability that the experimental result could have arisen by chance. By convention, if the p value is less than .05, meaning that the result could have arisen by chance less than 5% of the time, scientists accept the result as statistically significant. Of course, $p < .05$ is arbitrary, and it doesn't deal directly with the opposite case, the probability that the data you collected indicate a genuine effect, but the statistical test failed to detect it (a power analysis is required for this). In many studies, the probability of failing to detect an effect, when it exists, can soar to 80% (Schmidt 1996). An additional problem with a criterion of 5% is that a researcher who measures 20 different effects is likely to measure one as significant by chance, even if no significant effect actually exists.

Statistical significance tests, such as the analysis of variance (ANOVA), the f-test, chi-square test, and t-test, are methods to determine the probability that observed values in an experiment differ only as a result of measurement errors. For details about how to choose and conduct the appropriate tests, or to learn more about the theory behind them, consult a statistics textbook (e.g., Daniel 1990; Glenberg 1988; Hayes 1988).

Alternatives to Classical Significance Testing Because of problems with traditional significance testing, there is a movement, at the vanguard of applied statistics and psychology, to move away from "p value" tests and to rely on alternative methods, such as Bayesian inferencing, effect sizes, confidence intervals, and meta-analyses (refer to Cohen 1994; Hunter and Schmidt 1990; Schmidt 1996). Yet many people persist in clinging to the belief that the most important thing to do with experimental data is to test them for statistical significance. There is great pressure from peer-reviewed journals to perform significance tests, because so many people were taught to use them. The fact is, the whole point of significance testing is to determine whether a result is repeatable when one doesn't have the resources to repeat an experiment.

Let us return to the hypothetical example mentioned earlier, in which we examined the effect of music on study habits using a "within-subjects" design (each subject is in each condition). One possible outcome is that the difference in the mean test scores among groups was not significantly different by an analysis of variance (ANOVA). Yet suppose that, ignoring the means, every subject in the music-listening condition had a higher score than in the no-music condition. We are not interested in the size of the difference now, only in the direction of the difference. The null hypothesis predicts that the manipulation would have no effect at all, and that half of the subjects should show a difference in one direction and half in the other. The probability of all 10 subjects showing an effect in the same direction is $1/2^{10}$ or 0.0009, which is highly

significant. Ten out of 10 subjects indicates *repeatability.* The technique just described is called the *sign test*, because we are looking only at the arithmetic sign of the differences between groups (positive or negative).

Often, a good alternative to significance tests is estimates of *confidence intervals.* These determine with a given probability (e.g., 95%) the range of values within which the true population parameters lie. Another alternative is an analysis of *conditional probabilities.* That is, if you observe a difference between two groups on some measure, determine whether a subject's membership in one group or the other will improve your ability to predict his/her score on the dependent variable, compared with not knowing what group he/she was in (an example of this analysis is in Levitin 1994a). A good overview of these alternative statistical methods is contained in the paper by Schmidt (1996).

Aside from statistical analyses, in most studies you will want to compute the mean and standard deviation of your dependent variable. If you had distinct treatment groups, you will want to know the individual means and standard deviations for each group. If you had two continuous variables, you will probably want to compute the *correlation,* which is an index of how much one variable is related to the other. Always provide a table of means and standard deviations as part of your report.

6.8.2 Qualitative Analysis, or "How to Succeed in Statistics without Significance Testing"

If you have not had a course in statistics, you are probably at some advantage over anyone who has. Many people who have taken statistics courses rush to plug the numbers into a computer package to test for statistical significance. Unfortunately, students are not always perfectly clear on exactly what it is they are testing or why they are testing it.

The first thing one should do with experimental data is to graph them in a way that clarifies the relation between the data and the hypothesis. Forget about statistical significance testing—what does the pattern of data suggest? Graph everything you can think of—individual subject data, subject averages, averages across conditions—and see what patterns emerge. Roger Shepard has pointed out that the human brain is not very adept at scanning a table of numbers and picking out patterns, but is much better at picking out patterns in a visual display.

Depending on what you are studying, you might want to use a bar graph, a line graph, or a bivariate scatter plot. As a general rule, even though many of the popular graphing and spreadsheet packages will allow you to make pseudo-three-dimensional graphs, don't ever use three dimensions unless the third dimension actually represents a variable. Nothing is more confusing than a graph with extraneous information.

If you are making several graphs of the same data (such as individual subject graphs), make sure that each graph is the same size and that the axes are scaled identically from one graph to another, in order to facilitate comparison. Be sure all your axes are clearly labeled, and don't divide the axis numbers into units that aren't meaningful (for example, in a histogram with "number of subjects" on the ordinate, the scale shouldn't include half numbers because subjects come only in whole numbers).

Use a line graph if your variables are continuous. The lines connecting your plot points imply a continuous variable. Use a bar graph if the variables are categorical, so that you don't fool the reader into thinking that your observations were continuous. Use a bivariate scatter plot when you have two continuous variables, and you want to see how a change in one variable affects the other variable (such as how IQ and income might correlate). Do *not* use a bivariate scatterplot for categorical data. (For more information on good graph design, see Chambers et al. 1983; Cleveland 1994; Kosslyn 1994).

Once you have made all your graphs, look them over for interesting patterns and effects. Try to get a feel for what you have found, and understand how the data relate to your hypotheses and your experimental design. A well-formed graph can make a finding easy to understand and evaluate far better than a dry recitation of numbers and statistical tests can do.

Acknowledgments

This chapter benefited greatly from comments by Perry Cook, Lynn Gerow, Lewis R. Goldberg, John M. Kelley, and John Pierce. During the preparation of this chapter, I received direct support from an ONR graduate research fellowship (N-00014-89-J-3186), and indirect support from CCRMA and from an ONR Grant to M. I. Posner (N-00014-89-3013).

References

American Psychological Association. (1992). "Ethical Principles of Psychologists and Code of Conduct." *American Psychologist,* 47, 1597–1611.

American Psychological Association. (1994). *Publication Manual of the American Psychological Association.* Fourth edition. Washington, D.C.: American Psychological Association.

Butler, D., and W. D. Ward. (1988). "Effacing the Memory of Musical Pitch." *Music Perception,* 5 (3), 251–260.

Chambers, J. M., W. S. Cleveland, B. Kleiner, and P. A. Tukey. (1983). *Graphical Methods for Data Analysis.* New York: Chapman & Hall.

Cleveland, W. S. (1994). *The Elements of Graphing Data.* Revised edition. Summit, N.J.: Hobart Press.

Cohen, J. (1994). "The Earth Is Round ($p < .05$)." *American Psychologist,* 49, 997–1003.

Cozby, P. C. (1989). *Methods in Behavioral Research.* Fourth edition. Mountain View, Calif.: Mayfield Publishing Co.

Daniel, W. W. (1990). *Applied Nonparametric Statistics.* Second edition. Boston: PWS-Kent.

Deutsch, D. (1991). "The Tritone Paradox: An Influence of Language on Music Perception." *Music Perception,* 84, 335–347.

Deutsch, D. (1992). "The Tritone Paradox: Implications for the Representation and Communication of Pitch Structure." In M. R. Jones and S. Holleran, eds., *Cognitive Bases of Musical Communication.* Washington, D.C.: American Psychological Association.

Fisher, N. I. (1993). *Statistical Analysis of Circular Data.* Cambridge: Cambridge University Press.

Fletcher, H., and W. A. Munson. (1933). "Loudness, Its Definition, Measurement and Calculation." *Journal of the Acoustical Society of America,* 72, 82–108.

Glenberg, A. (1988). *Learning from Data: An Introduction to Statistical Reasoning.* San Diego: Harcourt, Brace, Jovanovich.

Hayes, W. (1988). *Statistics.* Fourth edition. New York: Holt, Rinehart and Winston.

Hempel, C. G. (1966). *Philosophy of Natural Science.* Englewood Cliffs, N.J.: Prentice-Hall.

Hunter, J. E., and F. L. Schmidt. (1990). *Methods of Meta-analysis: Correcting Error and Bias in Research Findings.* Newbury Park, Calif.: Sage.

Kirk, R. E. (1982). *Experimental Design: Procedures for the Behavioral Sciences.* Second edition. Pacific Grove, Calif.: Brooks/Cole.

Kosslyn, S. M. (1994). *Elements of Graph Design.* New York: Freeman.

Levitin, D. J. (1994a). "Absolute Memory for Musical Pitch: Evidence from the Production of Learned Melodies." *Perception & Psychophysics,* 56 (4), 414–423.

———. (1994b). *Problems in Applying the Kolmogorov-Smirnov Test: The Need for Circular Statistics in Psychology.* Technical Report #94-07. University of Oregon, Institute of Cognitive & Decision Sciences.

Schmidt, F. L. (1996). "Statistical Significance Testing and Cumulative Knowledge in Psychology: Implications for the Training of Researchers." *Psychological Methods,* VI (2): 115–129.

Shaughnessy, J. J., and E. B. Zechmeister. (1994). *Research Methods in Psychology.* Third edition. New York: McGraw-Hill.

Stern, A. W. (1993). "Natural Pitch and the A440 Scale." Stanford University, CCRMA. (Unpublished report).

Watson, J. B. (1967). *Behavior: An Introduction to Comparative Psychology.* New York: Holt, Rinehart and Winston. First published 1914.

Zar, J. H. (1984). *Biostatistical Analysis.* Second edition. Englewood Cliffs, N.J.: Prentice-Hall.

PART V

Perception

Chapter 7

Perception

Philip G. Zimbardo and Richard J. Gerrig

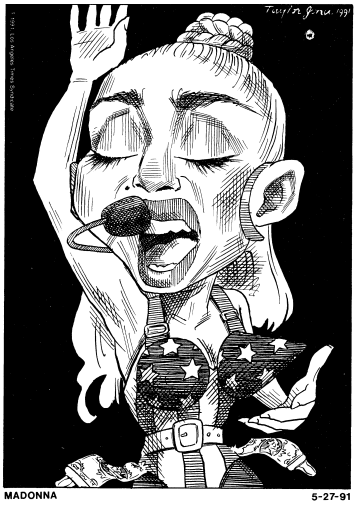
Who are the people in figure 7.1? If their fame has not been too fleeting, you should be able to recognize each of these individuals. But is this what they really look like? Probably not, at least on their good days. Your skill at identifying each of these caricatures suggests that your *perception* of the world relies on more than just the information arriving at your sensory receptors. Your ability to transform and interpret sensory information—your ability to have what you know interact with what you see—allows you to recognize Madonna, Oprah Winfrey, and Bill Clinton from these exaggerated portraits.

Your environment is filled with waves of light and sound, but that's not the way in which you experience the world. You don't "see" waves of light; you see a poster on the wall. You don't "hear" waves of sound; you hear music from a nearby radio. Sensation is what gets the show started, but something more is needed to make a stimulus meaningful and interesting and, most important, to make it possible for you to respond to it effectively. The processes of *perception* provide the extra layers of interpretation that enable you to navigate successfully through your environment.

We can offer a simple demonstration to help you think about the relationship between sensation and perception. Hold your hand as far as you can in front of your face. Now move it toward you. As you move your hand toward your eyes, it will take up more and more of your visual field. You may no longer be able to see the poster on the wall in back of your hand. How can your hand block out the poster? Has your hand gotten bigger? Has the poster gotten smaller? Your answer must be "Of course not!" This demonstration tells you something about the difference between sensation and perception. Your hand can block out the poster because, as it comes closer to your face, the hand projects an increasingly larger image on your retina. It is your perceptual processes that allow you to understand that despite the change in the size of the projection on your retina, your hand—and the poster behind it—do not change in actual size.

We might say that the role of perception is to make sense of sensation. Perceptual processes extract meaning from the continuously changing, often chaotic, sensory input from external energy sources and organize it into stable, orderly percepts. A *percept* is what is perceived—the phenomenological, or experienced, outcome of the process of perception. It is *not* a physical object or its image in a receptor but, rather, the psychological product of perceptual

From chapter 8 in *Psychology and Life*, 14th ed. (New York: HarperCollins, 1996), 258–302. Reprinted with permission.



MADONNA 5-27-91

OPRAH WINFREY 1-19-87

BILL CLINTON 11-16-92

Figure 7.1
What enables you to recognize these celebrities?

activity. Thus your percept of your hand remains stable over changes in the size of the image because your interpretation is governed by stable perceptual activities. Most of the time, sensing and perceiving occur so effortlessly, continuously, and automatically that you take them for granted. It is our goal in this chapter to allow you to understand and appreciate the processes that afford you a suitable account of the world, with such apparent ease. We begin with an overview of perceptual processes in the visual domain.

Sensing, Organizing, Identifying, and Recognizing

The term *perception,* in its broad usage, refers to the overall process of *apprehending* objects and events in the external environment—to sense them, understand them, identify and label them, and prepare to react to them. The process of perception is best understood when we divide it into three stages: sensation, perceptual organization, and identification/recognition of objects.

Sensation refers to conversion of physical energy into the neural codes recognized by the brain. Sensation provides a first-pass representation of the basic facts of the visual field. Your retinal cells are organized to emphasize edges and contrasts while reacting only weakly to unchanging, constant stimulation. Cells in your brain's cortex extract features and spatial frequency information from this retinal input.

Perceptual organization refers to the next stage, in which an internal representation of an object is formed and a percept of the external stimulus is developed. The representation provides a working description of the perceiver's external environment. Perceptual processes provide estimates of an object's likely size, shape, movement, distance, and orientation. Those estimates are based on mental computations that integrate your past knowledge with the present evidence received from your senses and with the stimulus within its perceptual context. Perception involves *synthesis* (integration and combination) of simple sensory features, such as colors, edges, and lines, into the percept of an object that can be recognized later. These mental activities most often occur swiftly and efficiently, without conscious awareness.

To understand the difference between these first two stages more clearly, consider the case study of Dr. Richard, whose brain damage left his sensation intact but altered his perceptual processes.

> Dr. Richard was a psychologist with considerable training and experience in introspection. This special skill enabled him to make a unique and valuable contribution to psychology. However, tragically, he suffered brain damage that altered his visual experience of the world. Fortunately, the damage did not affect the centers of his brain responsible for speech, so he was able to describe quite clearly his subsequent unusual visual experiences. In general terms, the brain damage seemed to have affected his ability to put sensory data together properly. For example, Dr. Richard reported that if he saw a complex object, such as a person, and there were several other people nearby in his visual field, he sometimes saw the different parts of the person as separate parts, not belonging together in a single form. He also had difficulty combining the sound and sight of the

same event. When someone was singing, he might see a mouth move and hear a song, but it was as if the sound had been dubbed with the wrong tape in a foreign movie.

To see the parts of an event as a whole, Dr. Richard needed some common factor to serve as "glue." For example, if the fragmented person moved, so that all parts went in the same direction, Dr. Richard would then perceive the parts reunited into a complete person. Even then, the perceptual "glue" would sometimes result in absurd configurations. Dr. Richard would frequently see objects of the same color, such as a banana, a lemon, and a canary, going together even if they were separated in space. People in crowds would seem to merge if they were wearing the same colored clothing. Dr. Richard's experiences of his environment were disjointed, fragmented, and bizarre—quite unlike what he had been used to before his problems began (Marcel, 1983).

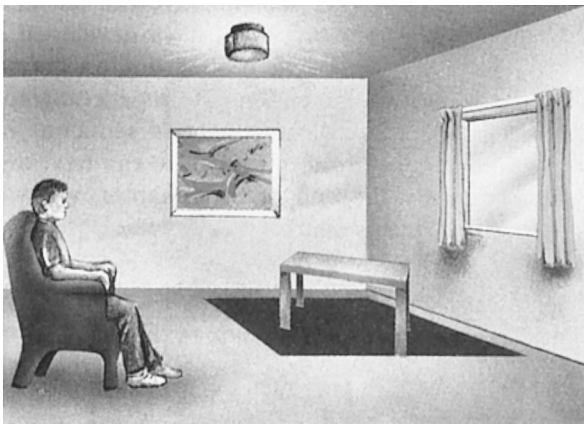
There was nothing wrong with Dr. Richard's eyes or with his ability to *analyze* the properties of stimulus objects—he saw the parts and qualities of objects accurately. Rather, his problem lay in synthesis—putting the bits and pieces of sensory information together properly to form a unified, coherent perception of a single event in the visual scene. His case makes salient the distinction between sensory and perceptual processes. It also serves to remind you that both sensory analysis and perceptual organization must be going on all the time even though you are unaware of the way they are working or even that they are happening.

Identification and recognition, the third stage in this sequence, assigns meaning to percepts. Circular objects "become" baseballs, coins, clocks, oranges, and moons; people may be identified as male or female, friend or foe, movie star or rock star. At this stage, the perceptual question "What does the object look like?" changes to a question of identification—"What is this object?"—and to a question of recognition—"What is the object's function?" To identify and recognize what something is, what it is called, and how best to respond to it involves higher level cognitive processes, which include your theories, memories, values, beliefs, and attitudes concerning the object.

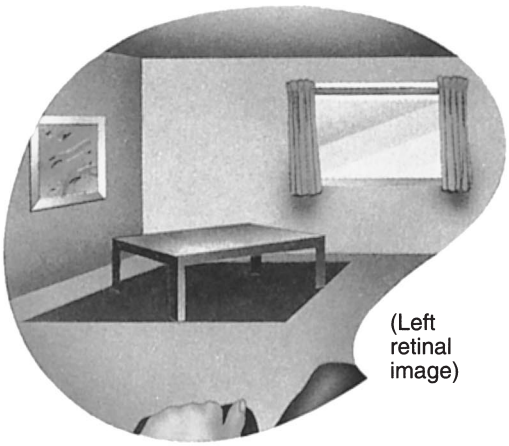
We have now given you a brief introduction to the stages of processing that enable you to arrive at a meaningful understanding of the perceptual world around you. We will devote the bulk of our attention here to aspects of perception beyond the initial transduction of physical energy. In everyday life, perception seems to be entirely effortless. We will try, beginning in the next section, to convince you that you actually do quite a bit of sophisticated processing, a lot of mental work, to arrive at this "illusion of ease."

The Proximal and Distal Stimulus

Imagine you are the person in figure 7.2, surveying a room from an easy chair. Some of the light reflected from the objects in the room enters your eyes and forms images on your retinas. Figure 7.2 shows what would appear to your left eye as you sat in the room. (The bump on the right is your nose, and the hand and knee at the bottom are your own.) How does this retinal image compare with the environment that produced it?



A. Physical object (distal stimulus)



B. Optical image (proximal stimulus)

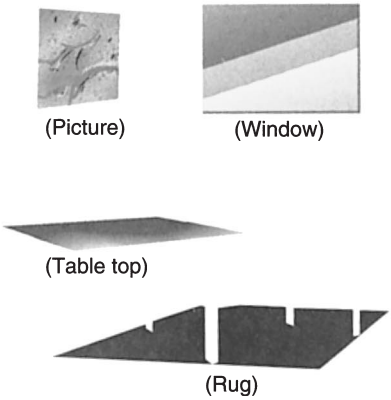


Figure 7.2
Interpreting retinal images.
A. Physical object (distal stimulus)
B. Optical image (proximal stimulus)

One very important difference is that the retinal image is *two-dimensional,* whereas the environment is *three-dimensional.* This difference has many consequences. For instance, compare the shapes of the physical objects in figure 7.2 with the shapes of their corresponding retinal images. The table, rug, window, and picture in the real-world scene are all rectangular, but only the image of the window actually produces a rectangle in your retinal image. The image of the picture is a trapezoid, the image of the table top is an irregular four-sided figure, and the image of the rug is actually three separate regions with more than 20 different sides! Here's our first perceptual puzzle: How do you manage to perceive all of these objects as simple, standard rectangles?

The situation is, however, even a bit more complicated. You can also notice that many parts of what you perceive in the room are not actually present in your retinal image. For instance, you perceive the vertical edge between the two walls as going all the way to the floor, but your retinal image of that edge stops at the table top. Similarly, in your retinal image parts of the rug are hidden behind the table; yet this does not keep you from correctly perceiving the rug as a single, unbroken rectangle. In fact, when you consider all the differences between the environmental objects and the images of them on your retina, you may be surprised that you perceive the scene as well as you do.

The differences between a physical object in the world and its optical image on your retina are so profound and important that psychologists distinguish carefully between them as two different stimuli for perception. The physical object in the world is called the *distal stimulus* (distant from the observer) and the optical image on the retina is called the *proximal stimulus* (proximate, or near, to the observer), as shown in figure 7.3.

The critical point of our discussion can now be restated more concisely: what you *perceive* corresponds to the *distal stimulus*—the "real" object in the environment—whereas the stimulus from which you must derive your information is the *proximal stimulus*—the image on the retina. The major computational task

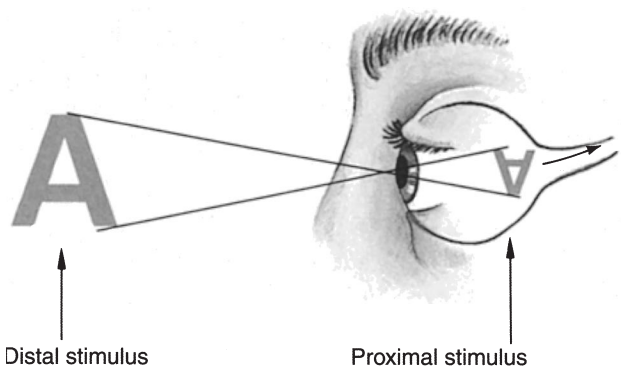


Figure 7.3
Distal and proximal stimulus. The distal stimulus is the pattern or external condition that is sensed and perceived. The proximal stimulus is the pattern of sensory activity that is determined by the distal stimulus. As illustrated here, the proximal stimulus may resemble the distal stimulus, but they are separate events.

of perception can be thought of as the process of determining the distal stimulus from information contained in the proximal stimulus. This is true across perceptual domains. For hearing, touch, taste, and so on, perception involves processes that use information in the proximal stimulus to tell you about properties of the distal stimulus.

To show you how the distal stimulus and proximal stimulus fit with the three stages in perceiving, let's examine one of the objects in the scene from figure 7.2: the picture hanging on the wall. In the sensory stage, this picture corresponds to a two-dimensional trapezoid in your retinal image; the top and bottom sides converge toward the right, and the left and right sides are different in length. This is the proximal stimulus. In the perceptual organization stage, you see this trapezoid as a rectangle turned away from you in three-dimensional space. You perceive the top and bottom sides as parallel, but receding into the distance toward the right; you perceive the left and right sides as equal in length. Your perceptual processes have developed a strong *hypothesis* about the physical properties of the distal stimulus; now it needs an identity. In the recognition stage, you identify this rectangular object as a picture. Figure 7.4 is a flowchart illustrating this sequence of events. The processes that take information from one stage to the next are shown as arrows between the boxes. By the end of this chapter, we will explain all the interactions represented in this figure.

Reality, Ambiguity, and Illusions

We have defined the task of perception as the identification of the distal stimulus from the proximal stimulus. Before we turn to some of the perceptual mechanisms that make this task successful, we want to discuss a bit more some other aspects of stimuli in the environment that make perception complex. Once again, you should look forward to learning how your perceptual processes deal with these complexities. We will discuss *ambiguous* stimuli and perceptual *illusions.*

Ambiguity A primary goal of perception is to get an accurate "fix" on the world. Survival depends on accurate perceptions of objects and events in your environment—Is that motion in the trees a tiger?—but the environment is not always easy to read. Take a look at the photo of black-and-white splotches in figure 7.5. What is it? Try to extract the stimulus figure from the background. Try to see a dalmatian taking a walk. The dog is hard to find because it blends with the background, so its boundaries are not clear. (Hint: the dog is on the right side of the figure, with its head pointed toward the center.) This figure is *ambiguous* in the sense that critical information is missing, elements are in unexpected relationships, and usual patterns are not apparent. *Ambiguity* is an important concept in understanding perception because it shows that a *single image* at the sensory level can result in *multiple interpretations* at the perceptual and identification levels.

Figure 7.6 shows three examples of ambiguous figures. Each example permits two unambiguous but conflicting interpretations. Look at each image until you can see the two alternative interpretations. Notice that once you have seen both of them, your perception flips back and forth between them as you look at the

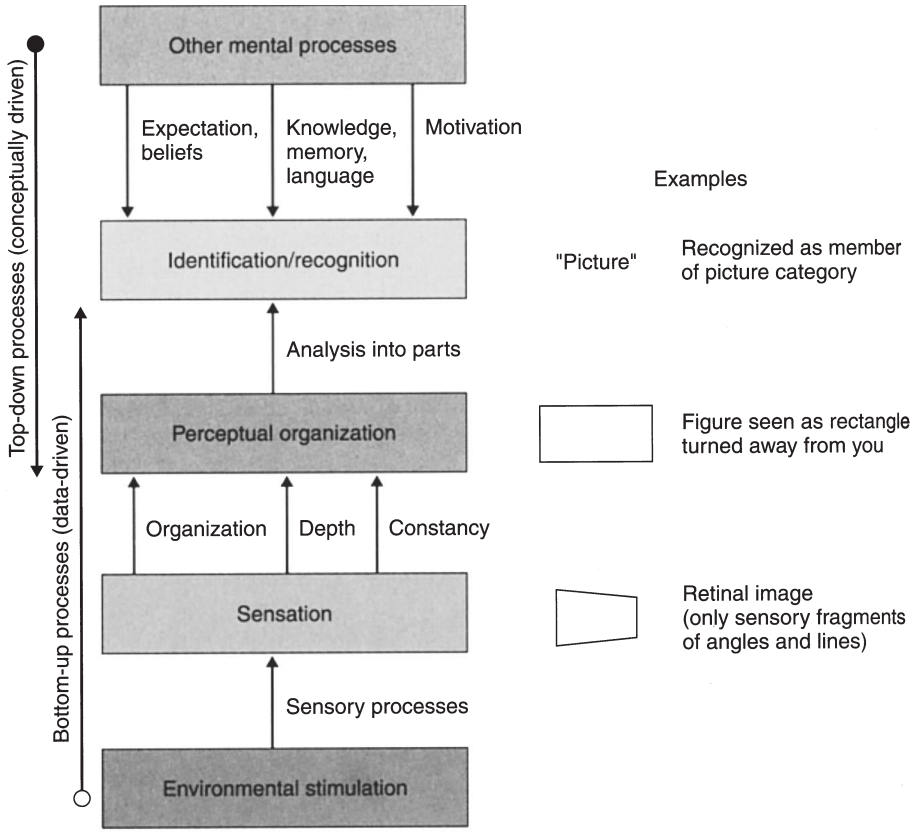


Figure 7.4
Sensation, perceptual organizing, and identification/recognition stages. The diagram outlines the processes that give rise to the transformation of incoming information at the stages of sensation, perceptual organization, and identification/recognition. Bottom-up processing occurs when the perceptual representation is derived from the information available in the sensory input. Top-down processing occurs when the perceptual representation is affected by an individual's prior knowledge, motivations, expectations, and other aspects of higher mental functioning.

ambiguous figure. This perceptual *instability* of ambiguous figures is one of their most important characteristics.

The vase/faces and the Necker cube are examples of ambiguity in the perceptual organization stage. You have two different perceptions of the same objects in the environment. The vase/faces can be seen as either a central white object on a black background or as two black objects with a white area between them. The Necker cube can be seen as a three-dimensional hollow cube either below you and angled to your left or above you and angled toward your right. With both vase and cube, the ambiguous alternatives are different physical arrangements of objects in three-dimensional space, both resulting from the same stimulus image.

The duck/rabbit figure is an example of ambiguity in the recognition stage. It is perceived as the same physical shape in both interpretations. The ambiguity

Figure 7.5
Ambiguous picture.

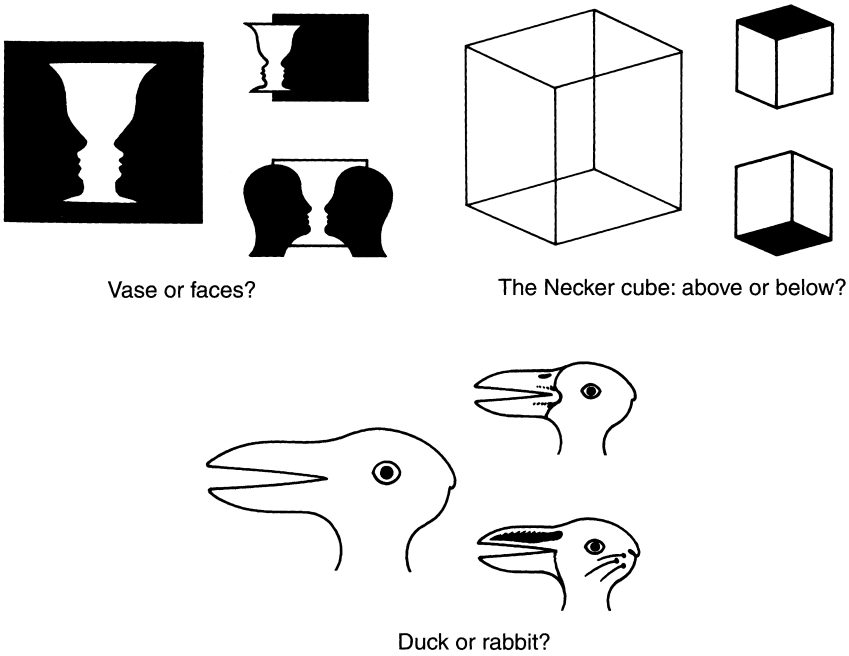


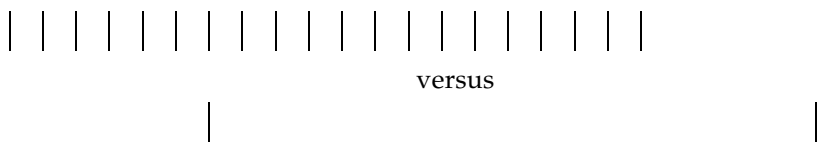
Figure 7.6
Perceptual ambiguities.

arises in determining the kind of object it represents and in how best to classify it, given the mixed set of information available.

One of the most fundamental properties of normal human perception is the tendency to transform ambiguity and uncertainty about the environment into a clear interpretation that you can act upon with confidence. In a world filled with variability and change, your perceptual system must meet the challenges of discovering invariance and stability.

Illusions Ambiguous stimuli present your perceptual systems with the challenge of recognizing one unique figure out of several possibilities. One or another interpretation of the stimulus is correct or incorrect with respect to a particular context. When your perceptual systems actually deceive you into experiencing a stimulus pattern in a manner that is demonstrably incorrect, you are experiencing an *illusion*. The word *illusion* shares the same root as *ludicrous*—both stem from the Latin *illudere,* which means "to mock at." Illusions are shared by most people in the same perceptual situation because of shared physiology in sensory systems and overlapping experiences of the world. (This sets illusions apart from *hallucinations.* Hallucinations are nonshared perceptual distortions that individuals experience as a result of unusual physical or mental states.) Examine the classic illusions in figure 7.7. Although it is most convenient for us to present you with visual illusions, illusions also exist abundantly in other sensory modalities such as hearing (Bregman, 1981; Shepard & Jordan, 1984) and taste (Todrank & Bartoshuk, 1991).

Since the first scientific analysis of illusions was published by J. J. Oppel in 1854–1855, thousands of articles have been written about illusions in nature, sensation, perception, and art. Oppel's modest contribution to the study of illusions was a simple array of lines that appeared longer when divided into segments than when only its end lines were present:



Oppel called his work the study of *geometrical optical illusions*. Illusions point out the discrepancy between percept and reality. They can demonstrate the abstract conceptual distinctions between sensation, perceptual organization, and identification and can help you understand some fundamental properties of perception (Cohen & Girgus, 1973).

Let's examine an illusion that works at the sensation level: the *Hermann grid*, in figure 7.8. As you stare at the center of the grid, dark, fuzzy spots appear at the intersections of the white bars. How does that happen? The answer lies in something you read about in the last chapter—*lateral inhibition*. Assume the stimulus is registered by ganglion retinal cells, two of which have their receptive fields drawn in the lower corner of the grid. The receptive field at the center of the intersection has two white bars projecting through its surround, while the neighboring receptive field has only one. The cell at the center, therefore, receives more light and can respond at a lower level because of the greater lateral inhibition by the surround. Its reduced response shows up as a dark spot

in its center. Illusions at this level generally occur because the arrangement of a stimulus array sets off receptor processes in an unusual way that generates a distorted image.

Illusions in Reality Are illusions just peculiar arrangements of lines, colors, and shapes used by artists and psychologists to plague unsuspecting people? Hardly. Illusions are a basic part of your everyday life. They are an inescapable aspect of the subjective reality you construct. And even though you may recognize an illusion, it can continue to occur and fool you again and again.

Consider your day-to-day experience of your home planet, the earth. You've seen the sun "rise" and "set" even though you know that the sun is sitting out there in the center of the solar system as decisively as ever. You can appreciate why it was such an extraordinary feat of courage for Christopher Columbus and other voyagers to deny the obvious illusion that the earth was flat and sail off toward one of its apparent edges. Similarly, when a full moon is overhead, it seems to follow you wherever you go even though you know the moon isn't chasing you. What you are experiencing is an illusion created by the great distance of the moon from your eye. When they reach the earth, the moon's light rays are essentially parallel and perpendicular to your direction of travel, no matter where you go.

People can control illusions to achieve desired effects. Architects and interior designers use principles of perception to create objects in space that seem larger or smaller than they really are. A small apartment becomes more spacious when it is painted with light colors and sparsely furnished with low, small couches, chairs, and tables in the center of the room instead of against the walls. Psychologists working with NASA in the U.S. space program have researched the effects of environment on perception in order to design space capsules that have pleasant sensory qualities. Set and lighting directors of movies and theatrical productions purposely create illusions on film and on stage.

Despite all of these illusions—some more useful than others—you generally do pretty well getting around the environment. That is why researchers typically study illusions to help explain why perception ordinarily works so well. The illusions themselves suggest, however, that your perceptual systems cannot perfectly carry out the task of recovering the distal stimulus from the proximal stimulus.

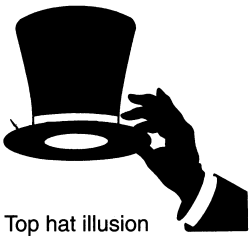
Approaches to the Study of Perception

You now are acquainted with some of the major questions of perception: How does the perceptual system recover the structure of the environment? How is ambiguity resolved? Why do illusions arise? Before we move on to answer these questions, we need to give you more of a background in the types of theories that have dominated research on perception.

Many of the differences between these theories can be captured by the distinction between *nature* and *nurture*. At issue is how much of a head start you have in dealing with the perceptual world by virtue of your possession of the human genotype. Do you, as a *nativist* might argue, come into the world with some types of innate knowledge or brain structures that aid your interpretation of the environment? Or do you, as an *empiricist* might assert, come into the

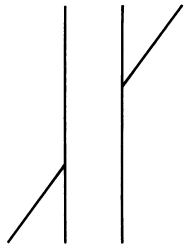
A. Use a ruler to answer each question.

Which is larger: the brim or the top hat?



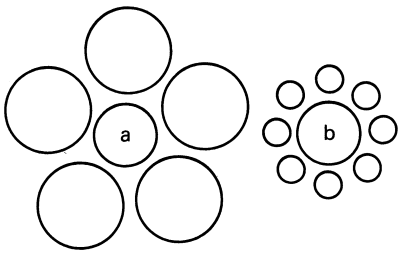
Top hat illusion

Is the diagonal line broken?



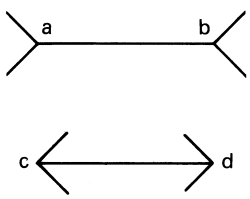
Poggendorf illusion

Which central circle is bigger?



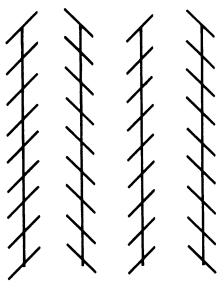
Ebbinghaus illusion

Which horizontal line is longer?



Müller-Lyer illusion

Are the vertical lines parallel?



Zöllner illusion

Figure 7.7
Six illusions to tease your brain.

B. Which of the boxes are the same size as the standard box? Which are definitely smaller or larger? Measure them to discover a powerful illusory effect.

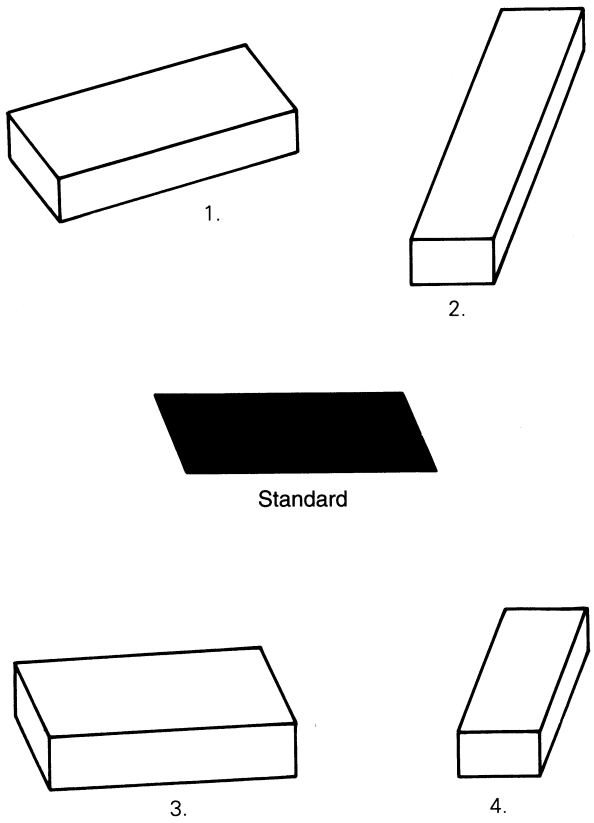


Figure 7.7 (continued)

world with a relatively blank slate, ready to learn what there is to learn about the perceptual world? Most modern theorists agree that your experience of the world consists of a combination of nature and nurture. We will see, however, that these theorists disagree on the size of the portions that make up this combination.

Helmholtz's Classical Theory In 1866, *Hermann von Helmholtz* argued for the importance of *experience*—or nurture—in perception. His theory emphasized the role of mental processes in interpreting the often ambiguous stimulus arrays that excite the nervous system. By using prior knowledge of the environment, an observer makes hypotheses, or inferences, about the way things really are. For instance, you would be likely to interpret your brief view of a four-legged creature moving through the woods as a dog rather than as a wolf. Perception is thus an *inductive* process, moving from specific images to

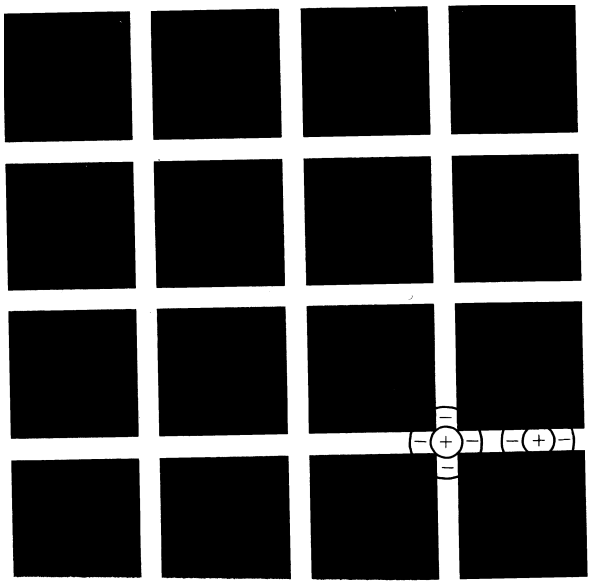


Figure 7.8
The Hermann grid. Two ganglion-cell receptive fields are projected on this grid; it is an example of an illusion at the sensory stage.

inferences (reasonable hunches) about the general class of objects or events that the images might represent. Since this process takes place out of your conscious awareness, Helmholtz termed it *unconscious inference.* Ordinarily, these inferential processes work well. However, perceptual illusions can result when unusual circumstances allow multiple interpretations of the same stimulus or favor an old, familiar interpretation when a new one is required.

Helmholtz's theory broke perception down into two stages. In the first, *analytic* stage, the sense organs analyze the physical world into fundamental sensations. In the second, *synthetic* stage, you integrate and synthesize these sensory elements into perceptions of objects and their properties. Helmholtz's theory proposes that you learn how to interpret sensations on the basis of your experience with the world. Your interpretations are, in effect, informed guesses about your perceptions.

The Gestalt Approach *Gestalt psychology,* founded in Germany in the second decade of the twentieth century, put greater emphasis on the role of innate structures—nature—in perceptual experience. The main exponents of Gestalt psychology, like *Kurt Koffka* (1935), *Wolfgang Köhler* (1947), and *Max Wertheimer* (1923), maintained that psychological phenomena could be understood only when viewed as organized, structured *wholes* and not when broken down into primitive perceptual elements. The term *Gestalt* roughly means "form," "whole," "configuration," or "essence." Gestalt psychology challenged atomistic views of psychology by arguing that the whole is more than the sum of its parts. For example, when you listen to music, you perceive whole melodies even though they are composed of separate notes. Gestalt psychologists argued

that the holistic perception of the world arises because the cortex is organized to function that way. You organize sensory information the way you do because it is the most economical, simple way to organize the sensory input, given the structure and physiology of the brain. (Many of the examples of perceptual organization we will discuss in a later section were originated by the Gestaltists.)

Gibson's Ecological Optics *James Gibson* (1966, 1979) proposed a very influential nativist approach to perception. Instead of trying to understand perception as a result of an organism's structure, Gibson suggested that it could be better understood through an analysis of the immediately surrounding environment (or its ecology). As one writer put it, Gibson's approach was, "Ask not what's inside your head, but what your head's inside of" (Mace, 1977). In effect, Gibson's *theory of ecological optics* was concerned with the perceived stimuli rather than with the mechanisms by which you perceive the stimuli. This approach was a radical departure from all previous theories. Gibson's ideas emphasized perceiving as *active exploration* of the environment. When an observer is *moving* in the world, the pattern of stimulation on the retina is constantly changing over time as well as over space. The theory of ecological optics tried to specify the information about the environment that was available to the eyes of a moving observer. Theorists in Gibson's tradition agree that perceptual systems evolved in organisms who were active—seeking food, water, mates, and shelter—in a complex and changing environment (Gibson, 1979; Pittenger, 1988; Shaw & Turvey, 1981; Shepard, 1984).

According to Gibson, the answer to the question "How do you learn about your world?" is simple. You directly pick up information about the *invariant,* or stable, properties of the environment. There is no need to take raw sensations into account or to look for higher level systems of perceptual inference—perception is direct. While the retinal size and shape of each environmental object changes, depending on the object's distance and on the viewing angle, these changes are not random. The changes are systematic, and certain properties of objects remain invariant under all such changes of viewing angles and viewing distances. Your visual system is tuned to detect such invariances because humans evolved in the environment in which perception of invariances was important for survival (Palmer, 1981).

Toward a Unified Theory of Perception These diverse theories can be unified to set the agenda for successful research on perception. You can recognize that the different perspectives contribute different insights to the three levels of analysis a theory of perception must address (Banks & Krajicek, 1991):

- *What are the physiological mechanisms involved in perception?* This topic has its history in work with animals, and has more recently been addressed using neuroimaging techniques (see Part 19). The information impinging on the sensory receptors is often ambiguous. Stimulus-driven, or bottom-up processing, works its way up the brain, while expectation-driven, or top-down processing, complements it.
- *What is the process of perceiving?* This question is usually tackled by researchers who follow in the tradition originated by Helmholtz and the

Gestaltists. Modern researchers often try to understand how sources of information are combined to arrive at a perceptual interpretation of the world. These researchers compare the process of perception to conceptual problem solving (Beck, 1982; Kanizsa, 1979; Pomerantz & Kubovy, 1986; Rock, 1983, 1986; Shepp & Ballisteros, 1989). We will see some of their insights in the remaining sections of this chapter.

• *What are the properties of the physical world that allow you to perceive?* This question makes contact with Gibson's theory. His central insight was that the world makes available certain types of information—and your perceptual apparatus is innately prepared to recover that information. Gibson's research made it clear that theories of perception must be constrained by accurate understandings of the environment in which people perceive.

We now begin our discussion of perceptual processes by considering what it means to select, or attend to, only a small subset of the information the world makes available.

Attentional Processes

We'd like you to take a moment now to find ten things in your environment that had not been, so far, in your immediate awareness. Had you noticed a spot on the wall? Had you noticed the ticking of a clock? If you start to examine your surroundings very carefully, you will discover that there are literally thousands of things on which you could focus your *attention.* Generally, the more closely you attend to some object or event in the environment, the more you can perceive and learn about it. That's why attention is an important topic in the study of perception: your focus of attention determines the types of information that will be most readily available to your perceptual processes. As you will now see, researchers have tried to understand what types of environmental stimuli require your attention and how attention contributes to your experience of those stimuli. We will start by considering how attention functions to selectively highlight objects and events in your environment.

Selective Attention

We began this section by asking that you try to find—to bring into attention—several things that had, up to that point, escaped your notice. This thought experiment illustrated an important function of attention: to select some part of the sensory input for further processing. Let us see how you make decisions about the subset of the world to which you will attend, and what consequences those decisions have for the information readily available to you.

Determining the Focus of Attention What forces determine the objects that become the focus of your attention? The answer to this question has two components, which we will call goal-directed selection and stimulus-driven capture (Yantis, 1993). *Goal-directed selection* reflects the choices that you make about the objects to which you'd like to attend, as a function of your own goals. You are probably already comfortable with the idea that you can explicitly choose objects for particular scrutiny. *Stimulus-driven capture* occurs when features of the stimuli—objects in the environment—themselves automatically *capture*

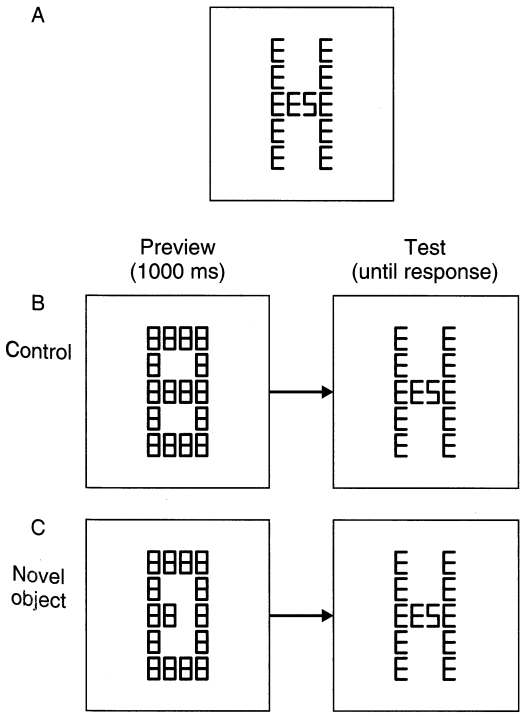


Figure 7.9
Stimulus-driven capture. How hard is it to recognize that the figure in (A) is an H? When the S fills a prior gap in the display (C), subjects find it more difficult to see that the overall figure is an H than they do in the control condition (B).

your attention, independent of your local goals as a perceiver. Research suggests, for example, that new objects in a perceptual display automatically capture attention.

Consider the figure shown in part A of figure 7.9. How hard do you think it would be for you to identify the overall, global figure as an H? The answer will depend on the extent to which you have to attend to the local letters that make up the global figure. Parts B and C of the figure show how researchers manipulated attention. In each condition of the experiment, subjects were given a preview display that consisted of a figure 8 made of 8s. In the control condition, the figure 8 was complete. But, as you can see, in the novel object condition, there was a gap in the figure. What will happen if the next display you see fills in that gap? The researchers predicted that the object filling the gap (the novel object) would capture your attention—you couldn't help looking at it. And if your attention is focused on the letter S, you should find it harder than you ordinarily would to say that the global letter is an H.

That is exactly the result the researchers obtained. If you compare the two test displays in figure 7.9, you'll see that they are identical. In each case an S helps to make up the global H. However, it was only in the case

> when the S appeared in a space that was previously unoccupied that subjects' performance—the speed with which they could name the global letter—was impaired (Hillstrom & Yantis, 1994).

You can recognize this phenomenon as stimulus-driven capture, because it works in the opposite direction of the perceiver's goals. Because, that is, the subjects would perform the task better if they ignored the small S, they must be unable to ignore it (since subjects almost always prefer to perform as well as possible on the tasks researchers assign them). The important general conclusion is that your perceptual system is organized so that your attention is automatically drawn to objects that are new to an environment.

The Fate of Unattended Information If you have selectively attended to some subset of a perceptual display—by virtue of your own goals or of properties of the stimuli—what is the fate of the information to which you did not attend? Imagine listening to a lecture while people on both sides of you are engaged in conversations. How are you able to keep track of the lecture? What do you notice about the conversations? Could anything appear in the content of one or the other conversation to divert your attention from the lecture?

This constellation of questions was first explored by *Donald Broadbent* (1958), who conceived of the mind as a *communications channel*—similar to a telephone line or a computer link—that actively processes and transmits information. Broadbent re-created the real-life situation of multiple sources of input in his laboratory with a technique called *dichotic listening*.

> A subject wearing earphones listens to two tape-recorded messages played at the same time—a different message is played into each ear. The subject is instructed to repeat only one of the two messages to the experimenter, while ignoring whatever is presented to the other ear. This procedure is called *shadowing* the attended message (see figure 7.10).

Subjects in shadowing experiments remember the attended message and do not remember the ignored message. Subjects usually do not even notice major alterations in the ignored message, such as changing the language from English to German or playing the tape backward. However, they do notice marked physical changes as, for example, when the pitch is raised substantially by changing the speaker's voice from male to female (Cherry, 1953). Thus gross physical features of the unattended message receive perceptual analysis, apparently below the level of consciousness, but most meaning does not get through.

According to Broadbent's theory, as a communications channel the mind has only *limited capacity* to carry out complete processing. This limit requires that attention strictly regulate the flow of information from sensory input to consciousness. Attention creates a bottleneck in the flow of information through the cognitive system, filtering out some information and allowing other information to continue. The *filter theory* of attention asserted that the selection occurs early on in the process, before the input's meaning is accessed.

The strongest form of filter theory was challenged when it was discovered that some subjects were perceiving things they would not have been able to if

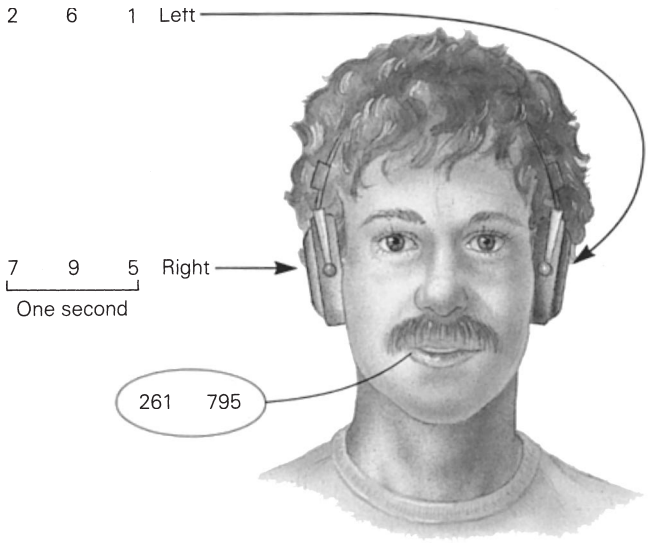


Figure 7.10
Dichotic listening task. A subject hears different digits presented simultaneously to each ear: 2 (left), 7 (right), 6 (left), 9 (right), 1 (left), and 5 (right). He reports hearing the correct sets—261 and 795. However, when instructed to attend only to the right-ear input, the subject reports hearing only 795.

attention had been totally filtering all ignored material. In dichotic listening tasks, subjects sometimes noticed their own names and other personally meaningful information contained in the message they were instructed to ignore (Cherry, 1953). When a story being shadowed in one ear was switched to the unattended ear and replaced by a new story, some subjects continued to report words from the original story, even though it was now entering the supposedly ignored ear. The subjects did so even though they had been accurately following the instruction about which ear to shadow (Treisman, 1960). Apparently, subjects were intrigued by the meaning and continuity of the particular message they had been shadowing, which momentarily distracted them from the attended channel. Some meaningful analysis of the ignored channel must have been taking place—otherwise, subjects would not have known that the message on that channel was the continuation of the message they had been shadowing. Therefore, attention does not function as an absolute filter. But then how does it function?

Research now suggests, in fact, that unattended objects are sufficiently processed by your perceptual system so that those objects become less available for later use (Tipper et al., 1991; Treisman, 1992).

> Look at figure 7.11. Try to read the black letters in each column. Disregard the overlapping gray letters. Did you notice that one of the columns is harder to read? Which one? Now look carefully at the gray letters. In the first column, there is no relationship between the gray letters and the black letters. However, in the second list, beginning with the second black

Figure 7.11
A test of your attentional mechanism. First, read aloud the black letters in Column one as quickly as possible, disregarding the gray. Next, quickly read the black letters in Column two, also disregarding the gray. Which took longer?

letter, each black letter is the same as the gray letter above it. A number of experiments show that subjects take longer to read the second list (Driver & Tipper, 1989; Tipper & Driver, 1988).

According to the authors of such experiments, subjects take longer to read the second column because they actually process the green letters unconsciously and have to inhibit or prevent themselves from responding to them. When, after having inhibited a particular letter, subjects are asked to respond to it, they are slowed down because they have to unblock or disinhibit the letter and make it available for response. For example, when you read the first black letter in the second row, you had to ignore, or inhibit, a gray H. The second black letter in the row happens to be the letter H. Thus, when you try to read the black H, you have to unblock, or disinhibit, the letter H. Nothing similar to this happens when you read the first row of letters; the black letters in this row never appear as gray letters.

Phenomena like this one suggest that selective attention works in two ways. First, your internal representations of the stimuli on which you have focused attention become highlighted in memory. Second, your internal representations of the unattended stimuli are somewhat suppressed. You can see how these processes of highlighting and suppression will make the attended objects specifically prominent in your consciousness. You can also see why it's dangerous to let yourself become distracted from your immediate task or goal. If you fail to pay attention to some body of information—your professor's lecture, perhaps—you may find it extra hard to catch up later. Let's turn now to the role attention plays in allowing you to find and correctly identify objects in your environment.

Attention and Objects in the Environment

One of the main functions of attention is to help you find particular objects in a noisy visual environment. To get a sense of how this works, you can carry out a very simple experiment. Put your book down for a minute and look for two things: a red object and a red object in the shape of a circle. Did it seem to you that you could find a red object almost instantly—without having to look at each part of the room—while finding the red circle required you to look around the room object by object? You have just discovered the difference between *preattentive processing* and processing that requires attention. We will now expand on these differences.

Preattentive Processing and Guided Search Even though conscious memory and recognition of objects require attention, quite complex processing of information goes on without attention and without awareness. This earlier stage of processing is called *preattentive processing* because it operates on sensory inputs before you attend to them, as they first come into the brain from the sensory receptors. The simple demonstration in figure 7.12 gives you a rough idea of what can and cannot be processed without attention (adapted from Rock & Gutman, 1981). Your memory for the attended (red) shapes in the figure is much better than memory for the unattended shapes. However, you remember some basic features of the unattended shapes, such as their color and whether they were drawn continuously or had gaps. It is as though your visual system extracted some of the simple features of the unattended objects but never quite managed to put them together to form whole percepts.

Preattentive processing is quite skilled at finding objects in the environment that can be defined by single features (Treisman & Sato, 1990; Wolfe, 1992). Look at part A of figure 7.13. Can you find the white T? This is a comparable exercise to finding a red object in the room around you. Preattentive processing allows you to search the environment in *parallel* for a single salient feature. This means that you can search all locations in the display at the same time: as a product of this parallel search, your attention is directed to the one correct object.

Now consider part B of figure 7.13. Try, once again, to find the white T. Didn't it feel harder? In this case, your attentional system is not equipped to differentiate white T's from white L's in a parallel search. You can still use your capability for parallel search to ignore all the black T's, but you must then

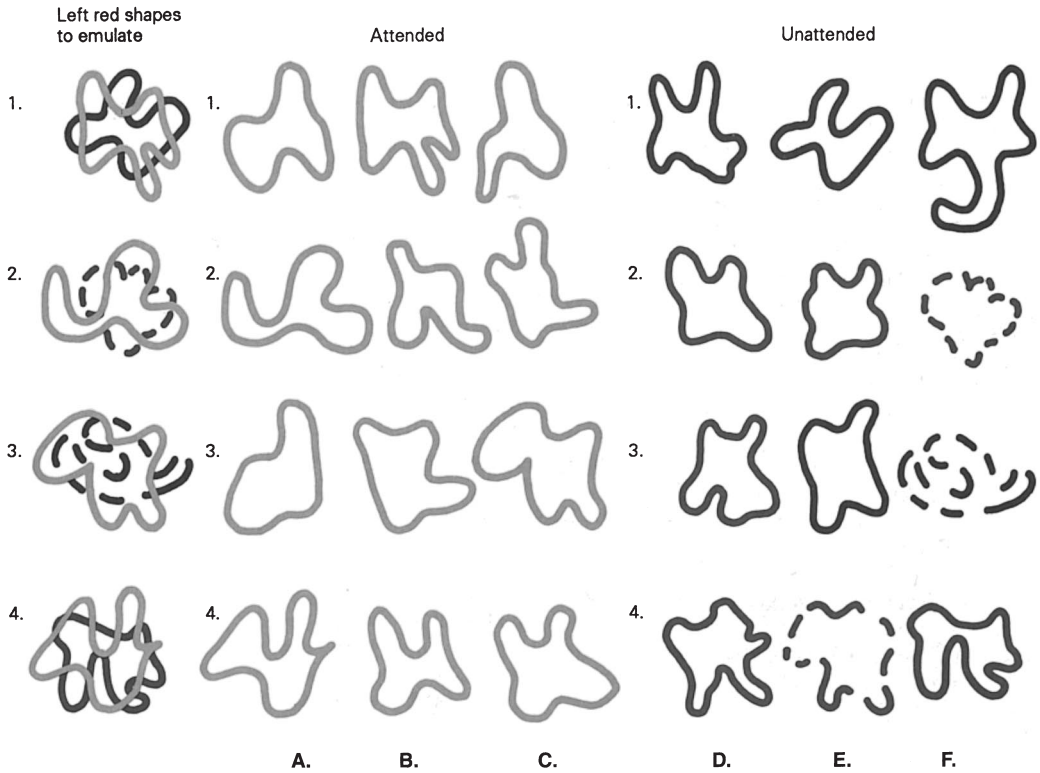


Figure 7.12
An example of overlapping figures. Cover the right part of the figure with a piece of paper. Look at the pictures of overlapping colored shapes on the left side of the figure. Try to attend to the red shapes only and rate them according to how appealing they seem to you. Next, cover the left side of the figure and uncover the right side. Now test your memory for the red (attended) figures and the blue and green (unattended) figures. Put a check mark next to each figure on the left you definitely recall seeing. How well do you remember the attended versus the unattended shapes?

consider each white symbol one by one, or *serially*. This experience is comparable to finding something in your environment that is both red and a circle. Preattentive processing allows you swiftly to find things that are red or things that are circles—preattentive processing allows a *guided search* of your environment (Wolfe, 1992). At that point, however, you need to attend to each object individually to determine whether it fits the *conjunction* of the two features round and red.

Researchers recognize the difference between a parallel and a serial search by determining how hard it is to find a target as a function of the number of distractors. Suppose we ask you to find a white T in a display with five black T's (as in part C of figure 7.13) versus a display with 34 black T's (as in part A). Because you can carry out this task in parallel, it will take you roughly the same amount of time to find the white T in each case. On the other hand, when you move from part B to part D of this figure, you can sense that you're much quicker to find the white T in D. You have to attend to each white element

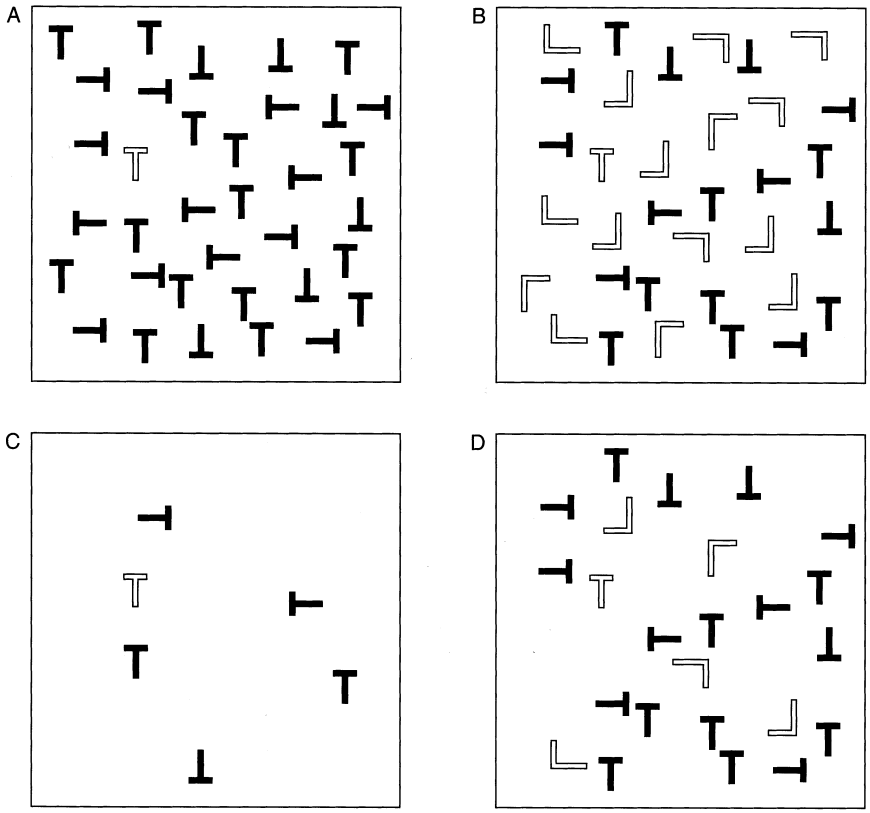


Figure 7.13
Attention and visual search.
(A) To find an object that differs on one salient feature, you can use parallel search.
(B) To find an object based on the conjunction of features, you must use serial search.
(C) Because parallel search is used, there is no difference in search time for this small array of distractors, as compared with the large array in part A.
(D) With serial search, the size of the array of distractors does make a difference. Search in D is faster than search in B.

serially, so each white element you look at (until you find the right one) adds a separate increment of time.

Researchers can use this logic to discover other aspects of the perceptual world that can be processed preattentively. Consider figure 7.14. In part A, try to find the yellow-and-blue item. In part B, try to find the yellow house with blue windows. Wasn't this second task much easier? Performance is much less affected by extra distractors when the two colors are organized into *parts* and *wholes* (Wolfe et al., 1994). Demonstrations of this sort suggest that preattentive processing provides you with relatively sophisticated assistance in finding objects in your environment.

Putting Features Together We have already seen that serially focused attention is often needed to find conjunctions of features. Researchers believe that, in general, putting the features of objects together into a complete percept requires

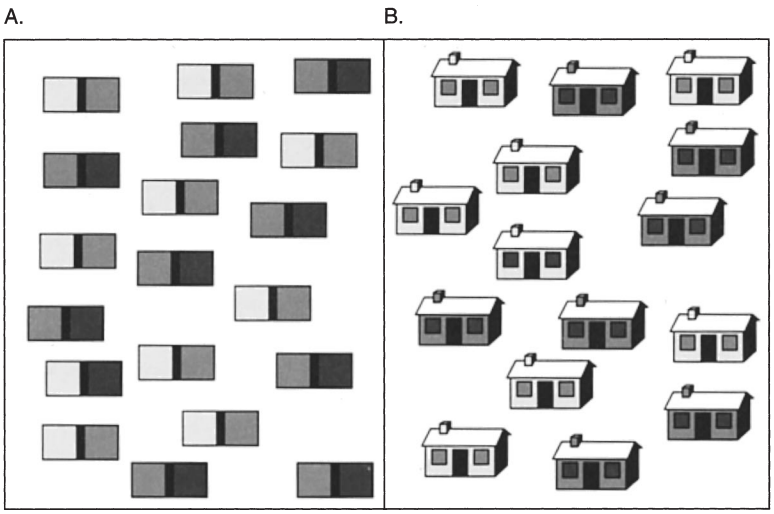


Figure 7.14
Search for the conjunction of two colors. (Yellow appears as light gray, red as medium gray, and blue as dark gray.)
(A) Find the yellow-and-blue item.
(B) Find the yellow house with blue windows.
(A) Search is very inefficient when the conjunction is between the colors of two parts of a target. (B) However, search is much easier when the conjunction is between the color of the whole item and the color of one its parts.

attention (Treisman, 1986, 1988; Treisman & Gelade, 1980). To demonstrate that attention is necessary to feature integration, researchers often divert or overload their subjects' attention. Under such circumstances, errors in feature combinations may occur, known as *illusory conjunctions.*

Researchers have produced illusory conjunctions by briefly flashing (for less than one-fifth of a second) three colored letters with digits on both sides of them.

5XOT7

The subjects' task is to report the digits first and then to report all of the color-letter combinations. On a third of the trials, subjects report seeing the wrong color-letter combination. For example, they report a yellow X instead of a blue X or a yellow O. They rarely make the mistake of reporting any colors or letters that were not present in the display, such as a red X or a blue Z.

Subjects were also likely to report that they saw a dollar sign ($) in the briefly flashed display containing S's and line segments shown in figure 7.15. The same effect was obtained even when the display contained S's and triangles. This result demonstrates that the subjects did not combine the lines of the triangles right away; the lines were floating unattached at some stage of perceptual processing, and one of the lines could be borrowed by the visual system to form the vertical bar in the dollar sign (Treisman & Gelade, 1980).

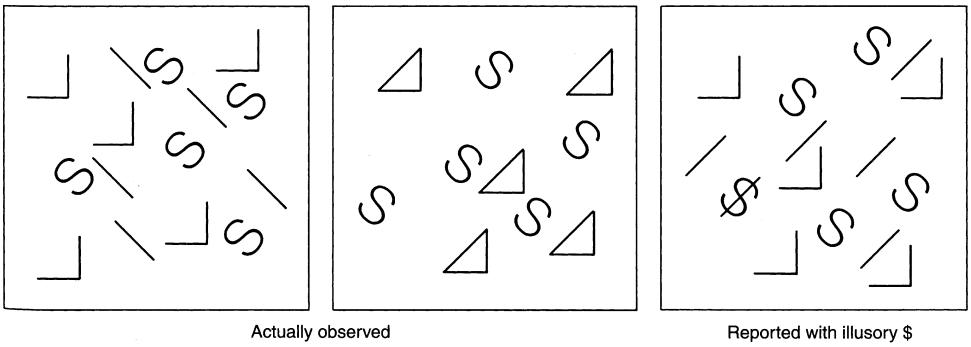


Figure 7.15
Combinations of features.

These results suggest that preattentive processing may allow perceivers to get individual features correct but, without focused attention, they are at risk for creating illusory conjunctions.

Illusory conjunctions also arise with more naturalistic stimuli. In one study, researchers used a slide projector to present subjects for 10 seconds with drawings of faces (Reinitz et al., 1994). Half of the subjects were put in a situation of *divided attention:* they were asked to count dots that appeared superimposed on the slide of each face. Later, both groups of subjects were asked to look at another series of slides and determine which of the faces they had seen before and which were new. The subjects in the divided attention condition were successful at recognizing the individual features of the faces—but they were inattentive to recombinations of those features. Thus, if a "new" face had the eyes from one "old" face and the mouth from another, they were as likely to say "old" as if the relations between the features had stayed intact. This result suggests that extracting facial features requires little or no attention, whereas extracting relationships between features *does* require attention. As a consequence, subjects who suffered from divided attention could remember what features they had seen but not which whole faces they belonged to!

If you make so many mistakes when putting the features together without attention in the laboratory, why don't you notice mistakes of this type when your attention is diverted or overloaded in the real world? Part of the answer is that you just might notice such mistakes if you start to look for them. It is common, for example, for eyewitnesses to give different accounts of the way the features of a crime situation combined to make the whole. Two witnesses might agree that *someone* was brandishing a gun but disagree on which of a team of bank robbers it was. Another part of the answer is provided by a leading researcher on attention, *Anne Treisman.* Treisman argues that most stimuli you process are familiar and sufficiently different from one another so that there are a limited number of sensible ways to combine their various features. Even when you have not attended as carefully as necessary for accurate integration of features, your knowledge of familiar perceptual stimuli allows you to guess how their features ought to be combined. These guesses, or perceptual hypotheses, are usually correct, which means that you construct some of your

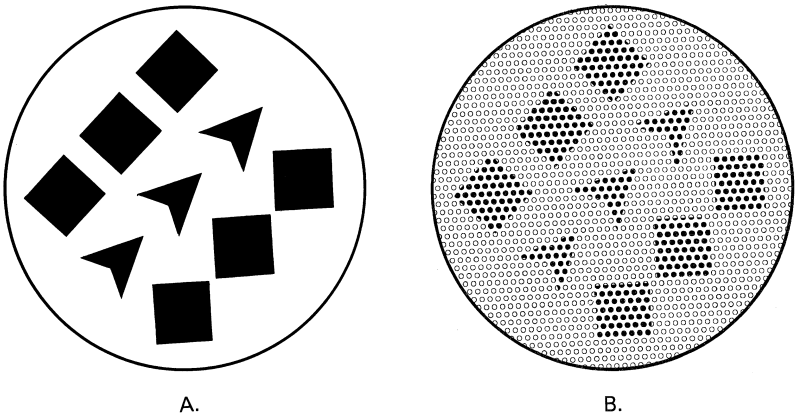


Figure 7.16
Percept of a two-dimensional geometric design. What is your percept of the geometrical design in A? B represents the mosaic pattern that stimulus A makes on your retina.

percepts by combining preattentive perception of single stimulus features with memory for familiar, similar whole figures.

We are now ready to make the transition from attention to individual features to the perception of whole objects and scenes.

Organizational Processes in Perception

Imagine how confusing the world would be if you were unable to put together and organize the information available from the output of your millions of retinal receptors. You would experience a kaleidoscope of disconnected bits of color moving and swirling before your eyes. The processes that put sensory information together to give you the perception of coherence are referred to collectively as processes of *perceptual organization.* You have seen that what a person experiences as a result of such perceptual processing is called a *percept.*

For example, your percept of the two-dimensional geometric design in part A of figure 7.16 is probably three diagonal rows of figures, the first being composed of squares, the second of arrowheads, and the third of diamonds. (We will discuss part B in a moment.) This probably seems unremarkable—but we have suggested in this chapter that all the seemingly effortless aspects of perception are made easy by sophisticated processing. Many of the organizational processes we will be discussing in this section were first described by Gestalt theorists who argued that what you perceive depends on laws of organization, or simple rules by which you perceive shapes and forms.

Region Segregation

Consider your initial sensory response to figure 7.16. Because your retina is composed of many separate receptors, your eye responds to this stimulus pattern with a mosaic of millions of independent neural responses coding the amount of light falling on tiny areas of your retina (see part B of figure 7.16). The first task of perceptual organization is to find coherent regions within this

Figure 7.17
Subjective contours that fit the angles of your mind.

mosaic of responses. In other words, your perceptual system must combine the outputs of the separate receptors into appropriate larger units. The primary information for this region-segregating process comes from color and texture. An abrupt change in color (hue, saturation, or brightness) signifies the presence of a boundary between two regions. Abrupt changes in texture can also mark boundaries between visibly different regions.

Researchers now believe that the feature-detector cells in the visual cortex, discovered by Hubel and Wiesel, are involved in these region-segregating processes (Marr, 1982). Some cells have elongated receptive fields that are ideally suited for detecting boundaries between regions that differ in color. Others have receptive fields that seem to detect bars or lines—of the sort that occur in grassy fields, wood grains, and woven fabrics. These cortical line-detector cells may be responsible for your ability to discriminate between regions with different textures (Beck, 1972, 1982; Julesz, 1981a, b).

Figure, Ground, and Closure

As a result of region segregation, the stimulus in figure 7.16 has now been divided into ten regions: nine small dark ones and a single large light one. You can think of each of these regions as a part of a unified entity, such as nine separate pieces of glass combined in a stained-glass window. Another organizational process divides the regions into figures and background. A *figure* is seen as an objectlike region in the forefront, and *ground* is seen as the backdrop against which the figures stand out. In figure 7.16, you probably see the dark regions as figures and the light region as ground. However, you can also see this stimulus pattern differently by reversing figure and ground, much as you did with the ambiguous vase/faces drawing and the Escher art. To do this, try to see the white region as a large white sheet of paper that has nine holes cut in it through which you can see a black background.

The tendency to perceive a figure as being in *front* of a ground is very strong. In fact, you can even get this effect in a stimulus when the perceived figure doesn't actually exist! In the first image of figure 7.17, you probably perceive a fir tree set against a ground containing several gray circles on a white surface.

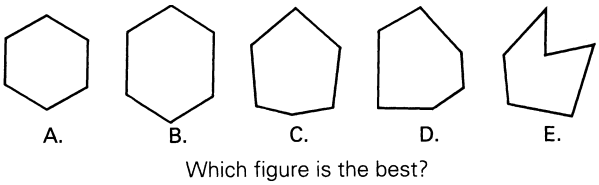


Figure 7.18
Figural goodness—1.

Notice, however, that there is no fir tree shape; the figure consists only of three solid gray figures and a base of lines. You see the illusory white triangle in front because the straight edges of the red shapes are aligned in a way that suggests a solid white triangle. The other image in figure 7.17 gives you the illusion of one complete triangle superimposed on another, although neither is really there.

In this example, there seem to be three levels of figure/ground organization: the white fir tree, the gray circles, and the larger white surface behind everything else. Notice that, perceptually, you divide the white area in the stimulus into two different regions: the white triangle and the white ground. Where this division occurs, you perceive illusory *subjective contours* that, in fact, exist not in the distal stimulus but only in your subjective experience.

Your perception of the white triangle in these figures also demonstrates another powerful organizing process: closure. *Closure* makes you see incomplete figures as complete. Though the stimulus gives you only the angles, your perceptual system supplies the edges in between that make the figure a complete fir tree. Closure processes account for your tendency to perceive stimuli as complete, balanced, and symmetrical, even when there are gaps, imbalance, or asymmetry.

Shape: Figural Goodness and Reference Frames

Once a given region has been segregated and selected as a figure against a ground, the boundaries must be further organized into specific *shapes.* You might think that this task would require nothing more than perceiving all the edges of a figure, but the Gestaltists showed that visual organization is more complex. If a whole shape were merely the sum of its edges, then all shapes having the same number of edges would be equally easy to perceive. In reality, organizational processes in shape perception are also sensitive to something the Gestaltists called *figural goodness,* a concept that includes perceived simplicity, symmetry, and regularity. Figure 7.18 shows several figures that exhibit a range of figural goodness even though each has the same number of sides. Do you agree that figure A is the "best" figure and figure E the "worst"?

Experiments have shown that good figures are more easily and accurately perceived, remembered, and described than bad ones (Garner, 1974). Such results suggest that shapes of good figures can be coded more rapidly and economically by the visual system. In fact, the visual system sometimes tends to see a single bad figure as being composed of two overlapping good ones, as shown in figure 7.19.

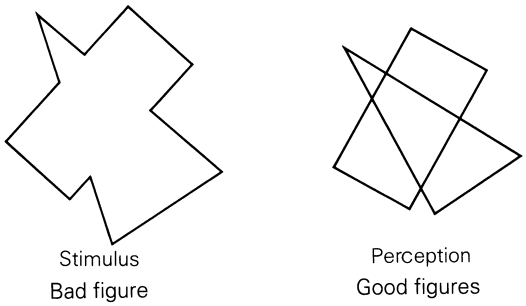


Figure 7.19
Figural goodness—2.

Your perceptual system also relies on *reference frames* to identify a figure's shape. Consider figure 7.20. If you saw the left-hand image in A by itself, it would resemble a diamond, whereas the right-hand image would resemble a square. When you see these images as parts of diagonal rows, as shown in B, the shapes reverse: the line composed of diamonds resembles a tilted column of squares, and the line composed of squares resembles a tilted column of diamonds. The shapes look different because the orientation of each image is seen in relation to the reference frame established by the whole row (Palmer, 1984, 1989). In effect, you see the shapes of the images as you would if the rows were vertical instead of diagonal (turn the book 45 degrees clockwise to see this phenomenon).

There are other ways to establish a contextual reference frame that has the same effect. These same images appear inside rectangular frames tilted 45 degrees in C of figure 7.20. If you cover the frames, the left image resembles a diamond and the right one a square. When you uncover the frames, the left one changes into a square and the right one into a diamond.

Principles of Perceptual Grouping

In figure 7.16, you perceived the nine figural regions as being grouped together in three distinct rows, each composed of three identical shapes placed along a diagonal line. How does your visual system accomplish this *perceptual grouping,* and what factors control it?

The problem of grouping was first studied extensively by Gestalt psychologist Max Wertheimer (1923). Wertheimer presented subjects with arrays of simple geometric figures. By varying a single factor and observing how it affected the way people perceived the structure of the array, he was able to formulate a set of laws of grouping. Several of these laws are illustrated in figure 7.21. In section A, there is an array of equally spaced circles that is ambiguous in its grouping—you can see it equally well as either rows or columns of dots. However, when the spacing is changed slightly so that the horizontal distances between adjacent dots are less than the vertical distances, as shown in B, you see the array unambiguously as organized into horizontal rows; when the spacing is changed so that the vertical distances are less, as shown in C, you see the array as organized into vertical columns. Together, these three groupings illustrate Wertheimer's *law of proximity:* all else being equal, the nearest (most

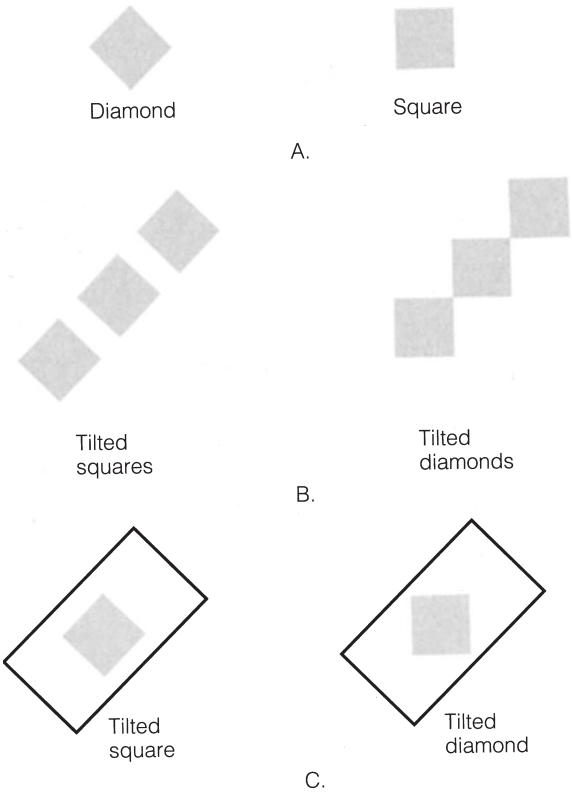


Figure 7.20
Reference frames.

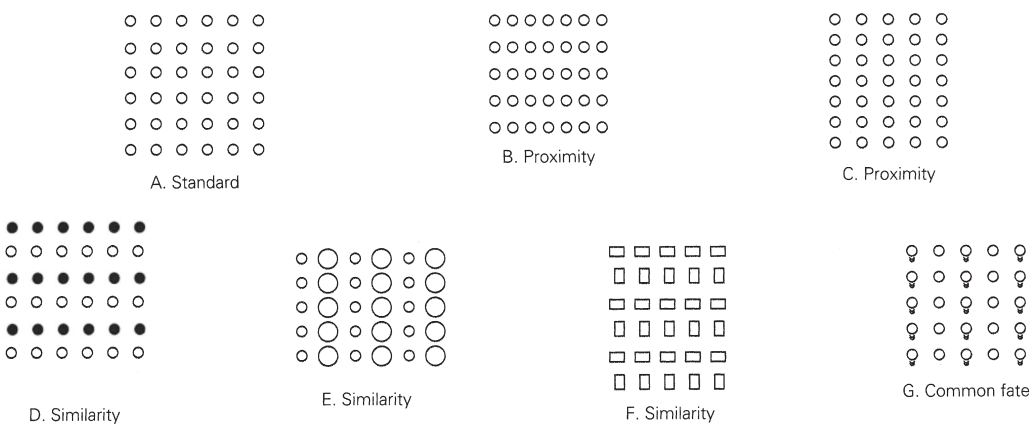


Figure 7.21
Grouping phenomena. We perceive each array from B through G as being organized in a particular way, according to different Gestalt principles of grouping.

proximal) elements are grouped together. The Gestaltists interpreted such results to mean that the whole stimulus pattern is somehow determining the organization of its own parts; in other words, the *whole percept* is different from the mere collection of its *parts.*

In D, the color of the dots instead of their spacing has been varied. Although there is equal spacing between the dots, your visual system automatically organizes this stimulus into rows because of their *similar color.* You see the dots in E as being organized into columns because of *similar size,* and you see the dots in F as being organized into rows because of *similar shape* and *orientation.* These grouping effects can be summarized by the *law of similarity:* all else being equal, the most similar elements are grouped together.

When elements in the visual field are moving, similarity of motion also produces a powerful grouping. The *law of common fate* states that, all else being equal, elements moving in the same direction and at the same rate are grouped together. If the dots in every other column of G were moving upward, as indicated by the blurring, you would group the image into columns because of their similarity in motion. You get this effect at a ballet when several dancers move in a pattern different from the others. Remember Dr. Richard's observation that an object in his visual field became organized properly when it moved as a whole. His experience was evidence of the powerful organizing effect of common fate.

Is there a more general way of stating the various grouping laws we have just discussed? We have mentioned the law of proximity, the law of similarity, the law of common fate, and the law of symmetry, or figural goodness. Gestalt psychologists believed that all of these laws are just particular examples of a general principle, the *law of pragnanz* (*pragnanz* translates roughly to "good figure"): you perceive the simplest organization that fits the stimulus pattern.

Spatial and Temporal Integration

All the Gestalt laws we have presented to you so far should have convinced you that a lot of perception consists of putting the pieces of your world together in the "right way." Often, however, you can't perceive an entire scene in one glance, or *fixation* (recall our discussion of attention). What you perceive at a given time is often a restricted glimpse of a large visual world extending in all directions to unseen areas of the environment. What may surprise you is that your visual system does not work very hard to create a moment-by-moment, integrated picture of the environment. Research suggests that your visual memory for each fixation on the world does not preserve precise details (Irwin, 1991). Why is that so? Part of the answer might be that the world itself is generally a stable source of information (O'Regan, 1992). It is simply unnecessary to commit to memory information that remains steadily available in the external environment.

One interesting consequence of the way you treat the information from different fixations is that you are taken in by illusions called "impossible" objects, such as those in figure 7.22. For example, each fixation of corners and sides provides an interpretation that is consistent with an object that seems to be a three-dimensional triangle (image A); but when you try to integrate them into a

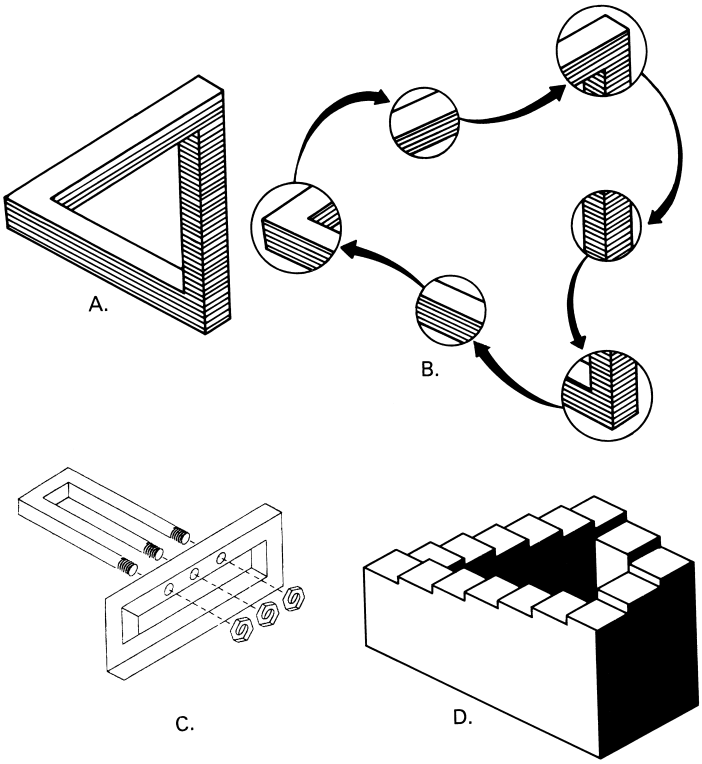


Figure 7.22
Impossible figures.

coherent whole, the pieces just don't fit together properly (image B). Image C has two arms that somehow turn into three prongs right before your vigilant gaze, and the perpetual staircase in image D forever ascends or descends.

Motion Perception

One type of perception that does require you to compare across different glimpses of the world is motion perception. Consider the two images given in figure 7.23. Suppose that this individual has stood still while you have walked toward him. The size of his image on your retina has expanded as you have drawn near. The rate at which this image has expanded gives you a sense of how quickly you have been approaching (Gibson, 1979). You use this type of information to navigate effectively in your world.

Suppose, however, you are still but other objects are in motion. The perception of motion, like the perception of shape and orientation, often depends on a reference frame. If you sit in a darkened room and fixate on a stationary spot of light inside a lighted rectangle that is moving very slowly back and forth, you will perceive instead a *moving* dot going back and forth within a *stationary* rectangle. This illusion, called *induced motion,* occurs even when your eyes are quite still and fixated on the dot. Your motion-detector cells are not firing at all in response to the stationary dot but presumably are firing in response to the

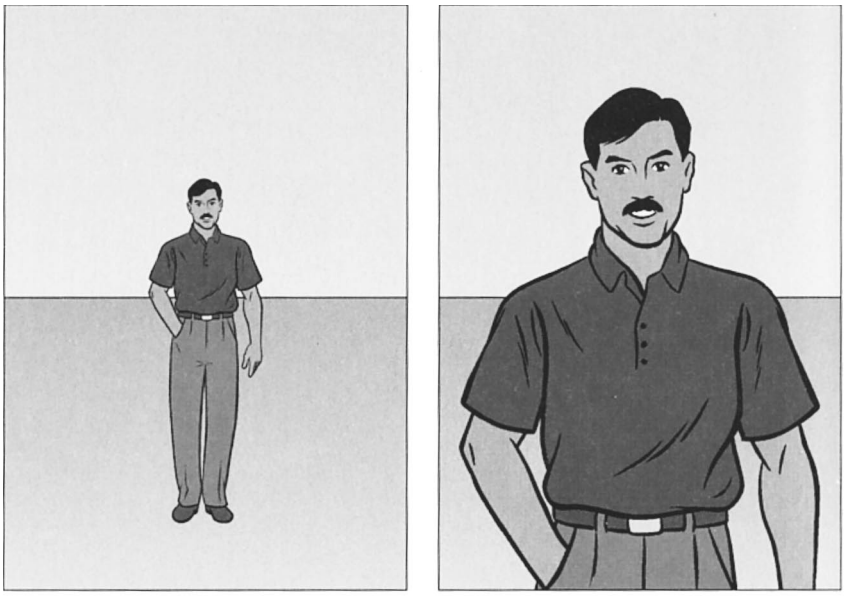


Figure 7.23
Approaching man. The size of an image expands on your retina as you draw nearer to the stimulus.

moving lines of the rectangle. To see the dot as moving requires some higher level of perceptual organization in which the dot and its supposed motion are perceived within the reference frame provided by the rectangle.

There seems to be a strong tendency for the visual system to take a larger, surrounding figure as the reference frame for a smaller figure inside it. You have probably experienced induced motion many times without knowing it. The moon (which is nearly stationary) frequently looks as if it is moving through a cloud, when, in fact, it is the cloud that is moving past the moon. The surrounding cloud induces perceived movement in the moon just as the rectangle does in the dot (Rock, 1983, 1986). Have you ever been in a train that started moving very slowly? Didn't it seem as if the pillars on the station platform or a stationary train next to you might be moving backward instead?

Another movement illusion that demonstrates the existence of higher level organizing processes for motion perception is called *apparent motion.* The simplest form of apparent motion, the *phi phenomenon,* occurs when two stationary spots of light in different positions in the visual field are turned on and off alternately at a rate of about 4 to 5 times per second. This effect occurs on outdoor advertising signs and in disco light displays. Even at this relatively slow rate of alternation, it appears that a single light is moving back and forth between the two spots. There are multiple ways to conceive of the path that leads from the location of the first dot to the location of the second dot. Yet human observers normally see only the simplest path, a straight line (Cutting & Proffitt, 1982; Shepard, 1984). This straight-line rule is violated, however, when subjects are shown alternating views of a human body in motion. Then the visual system fills in the paths of normal biological motion (Shiffrar, 1994).

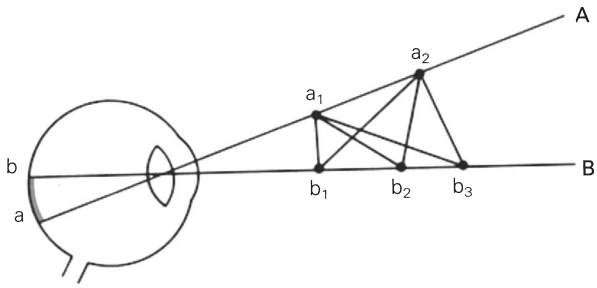


Figure 7.24
Depth ambiguity.

Depth Perception
Until now, we have considered only two-dimensional patterns on flat surfaces. Everyday perceiving, however, involves objects in three-dimensional space. Perceiving all three spatial dimensions is absolutely vital for you to approach what you want, such as interesting people and good food, and avoid what is dangerous, such as speeding cars and falling comets. This perception requires accurate information about *depth* (the distance from you to an object) as well as about its *direction* from you. Your ears can help in determining direction, but they are not much help in determining depth.

When you think about depth perception, keep in mind that the visual system must rely on retinal images that have only two spatial dimensions—vertical and horizontal. To illustrate the problem of having a 2-D retina doing a 3-D job, consider the situation shown in figure 7.24. When a spot of light stimulates the retina at point a, how do you know whether it came from position a_1 or a_2? In fact, it could have come from *anywhere* along line A, because light from any point on that line projects onto the same retinal cell. Similarly, all points on line B project onto the single retinal point b. To make matters worse, a straight line connecting any point on line A to any point on line B (a_1 to b_2 or a_2 to b_1, for example) would produce the same image on the retina. The net result is that the image on your retina is ambiguous in depth: it could have been produced by objects at any one of several different distances.

The two possible views of the Necker cube from figure 7.6 result from this ambiguity in depth. The fact that you can be fooled under certain circumstances shows that depth perception requires an *interpretation* of sensory input and that this interpretation can be wrong. (You already know this if you've ever swung at a tennis ball and come up only with air.) Your interpretation of depth relies on many different information sources about distance (often called *depth cues*)—among them binocular cues, motion cues, and pictorial cues.

Binocular and Motion Cues Have you ever wondered why you have two eyes instead of just one? The second eye is more than just a spare—it provides some of the best, most compelling information about depth. The two sources of binocular depth information are *binocular disparity* and *convergence*.

Because the eyes are about two to three inches apart horizontally, they receive slightly different views of the world. To convince yourself of this, try the

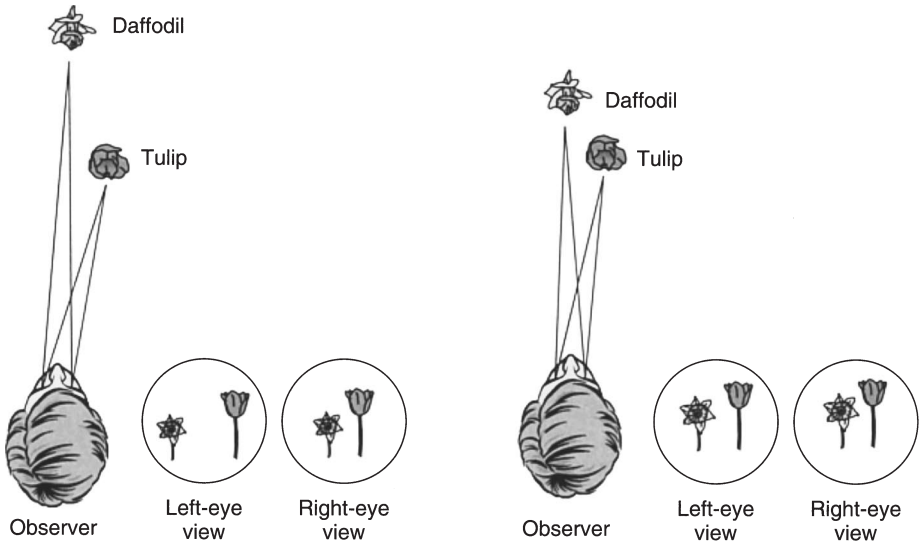


Figure 7.25
Retinal disparity. Retinal disparity increases with the distance, in depth, between two objects.

following experiment. First, close your left eye and use the right one to line up your two index fingers with some small object in the distance, holding one finger at arm's length and the other about a foot in front of your face. Now, keeping your fingers stationary, close your right eye and open the left one while continuing to fixate on the distant object. What happened to the position of your two fingers? The second eye does not see them lined up with the distant object because it gets a slightly different view.

This displacement between the horizontal positions of corresponding images in your two eyes is called *binocular disparity*. It provides depth information because the amount of disparity, or difference, depends on the relative distance of objects from you (see figure 7.25). For instance, when you switched eyes, the closer finger was displaced farther to the side than was the distant finger.

When you look at the world with both eyes open, most objects that you see stimulate different positions on your two retinas. If the disparity between corresponding images in the two retinas is small enough, the visual system is able to fuse them into a perception of a single object in depth. (However, if the images are too far apart, as when you cross your eyes, you actually see the double images.) When you stop to think about it, what your visual system does is pretty amazing: it takes two different retinal images, compares them for horizontal displacement of corresponding parts (binocular disparity), and produces a unitary perception of a single object in depth. In effect, the visual system interprets horizontal displacement between the two images as depth in the three-dimensional world.

Other binocular information about depth comes from *convergence*. The two eyes turn inward to some extent whenever they are fixated on an object (see figure 7.26). When the object is very close—a few inches in front of your face—

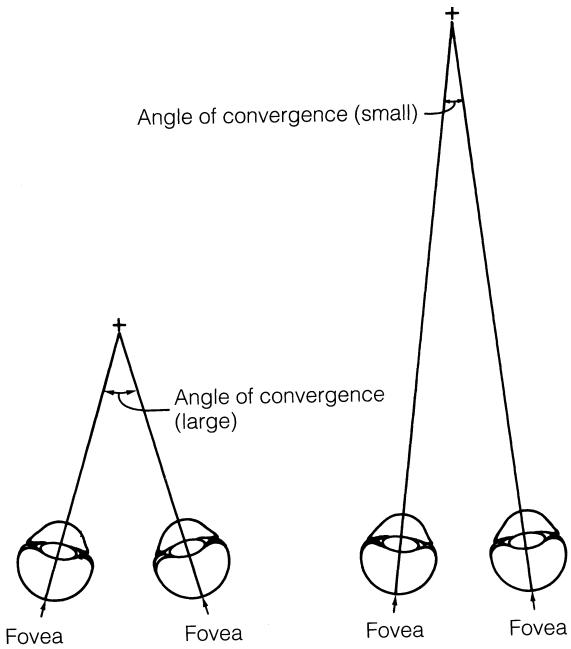


Figure 7.26
Convergence cues to depth.

the eyes must turn toward each other quite a bit for the same image to fall on both foveae. You can actually see the eyes converge if you watch a friend focus first on a distant object and then on one a foot or so away. Your brain uses information from your eye muscles to make judgments about depth. However, convergence information from the eye muscles is useful for depth perception only up to about 10 feet. At greater distances, the angular differences are too small to detect, because the eyes are nearly parallel when you fixate on a distant object.

To see how *motion* is another source for depth information, try the following demonstration. As you did before, close one eye and line up your two index fingers with some distant object. Then move your head to the side while fixating on the distant object and keeping your fingers still. As you move your head, you see both your fingers move, but the close finger seems to move farther and faster than the more distant one. The fixated object does not move at all. This source of information about depth is called *relative motion parallax.* Motion parallax provides information about depth because, as you move, the relative distances of objects in the world determine the amount and direction of their relative motion in your retinal image of the scene. Next time you are a passenger on a car trip, you should keep a watch out the window for motion parallax at work. Objects at a distance from the moving car will appear much more stationary than those closer to you.

Pictorial Cues But suppose you had vision in only one eye. Would you not be able to perceive depth? In fact, further information about depth is available

Figure 7.27
Interposition cues to depth. What are the visual cues that tell you whether or not this woman is behind the bars?

from just one eye. These sources are called *pictorial cues*, because they include the kinds of depth information found in pictures. Artists who create images in what appear to be three dimensions (on the two dimensions of a piece of paper or canvas) make skilled use of pictorial cues.

Interposition, or *occlusion*, arises when an opaque object blocks out part of a second object (see figure 7.27). Interposition gives you depth information indicating that the occluded object is farther away than the occluding one. Occluding surfaces also block out light, creating shadows that can be used as an additional source of depth information.

Three more sources of pictorial information are all related to the way light projects from a three-dimensional world onto a two-dimensional surface such as the retina: relative size, linear perspective, and texture gradients. *Relative size* involves a basic rule of light projection: objects of the same size at different distances project images of different sizes on the retina. The closest one projects the largest image and the farthest one the smallest image. This rule is called the *size/distance relation*. As you can see in figure 7.28, if you look at an array with identical objects, you interpret the smaller ones to be further away.

Linear perspective is a depth cue that also depends on the size/distance relation. When parallel lines (by definition separated along their lengths by the same distance) recede into the distance, they converge toward a point on the horizon in your retinal image (see figure 7.29). This important fact was discovered around 1400 by Italian Renaissance artists, who were then able to paint depth compellingly for the first time (Vasari, 1967). Prior to their discovery, artists had incorporated in their paintings information from interposition,

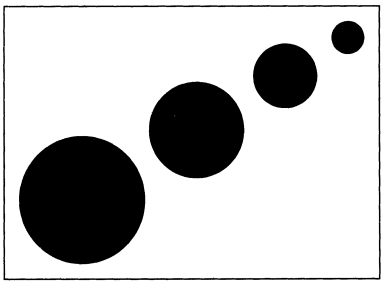


Figure 7.28
Relative size as a depth cue.

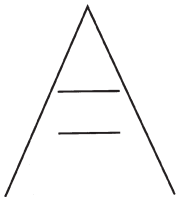


Figure 7.29
The Ponzo illusion. The converging lines add a dimension of depth, and, therefore, the distance cue makes the top line appear larger than the bottom line, even though they are actually the same length.

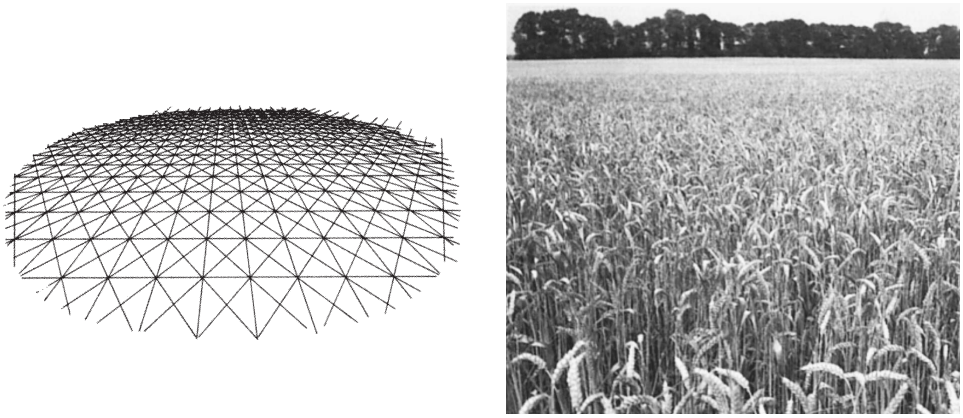


Figure 7.30
Examples of texture as a depth cue. The wheat field is a natural example of the way texture is used as a depth cue. Notice the way wheat slants. The geometric design uses the same principles.

shadows, and relative size, but they had been unable to depict realistic scenes that showed objects at various depths.

Your visual system's interpretation of converging lines gives rise to the *Ponzo illusion* (also shown in figure 7.29). The upper line looks longer because you interpret the converging sides according to linear perspective as parallel lines receding into the distance. In this context, you interpret the upper line as though it were farther away, so you see it as longer—a farther object would have to be longer than a nearer one for both to produce retinal images of the same size.

Texture gradients provide depth cues because the density of a texture becomes greater as a surface recedes in depth. The wheat field in figure 7.30 is an example of the way texture is used as a depth cue. You can think of this as another consequence of the size/distance relation. In this case, the units that make up the texture become smaller as they recede into the distance, and your visual system interprets this diminishing grain as greater distance in three-dimensional space. Gibson (1966, 1979) suggested that the relationship between texture and depth is one of the invariants available in the perceptual environment.

By now, it should be clear that there are many sources of depth information. Under normal viewing conditions, however, information from these sources comes together in a single, coherent three-dimensional interpretation of the environment. You experience depth, not the different cues to depth that existed in the proximal stimulus. In other words, your visual system uses cues like differential motion, interposition, and relative size automatically, without your conscious awareness, to make the complex computations that give you a perception of depth in the three-dimensional environment.

Perceptual Constancies

To help you discover another important property of visual perception, we are going to ask you to play a bit with your textbook. Put your book down on a table, then move your head closer to it so that it's just a few inches away. Then

move your head back to a normal reading distance. Although the book stimulated a much larger part of your retina when it was up close than when it was far away, didn't you perceive the book's size to remain the same? Now set the book upright and try tilting your head clockwise. When you do this, the image of the book rotates counterclockwise on your retina, but didn't you still perceive the book to be upright?

In general, you see the world as *invariant, constant*, and *stable* despite changes in the stimulation of your sensory receptors. Psychologists refer to this phenomenon as *perceptual constancy*. Roughly speaking, it means that you perceive the properties of the distal stimuli, which are usually constant, rather than the properties of proximal stimuli, which change every time you move your eyes or head. For survival, it is critical that you perceive constant and stable properties of objects in the world despite the enormous variations in the properties of the light patterns that stimulate your eyes. The critical task of perception is to discover *invariant* properties of your environment despite the *variations* in your retinal impressions of them. We will see how this works for size, shape, and orientation.

Size and Shape Constancy What determines your perception of the size of an object? In part, you perceive an object's actual size on the basis of the size of its retinal image. However, the demonstration with your book shows that the size of the retinal image depends on both the actual size of the book and its *distance* from the eye. As you now know, information about distance is available from a variety of depth cues. Your visual system combines that information with retinal information about image size to yield a perception of an object size that usually corresponds to the actual size of the distal stimulus. *Size constancy* refers to your ability to perceive the true size of an object despite variations in the size of its retinal image.

If the size of an object is perceived by taking distance cues into account, then you should be fooled about size whenever you are fooled about distance. One such illusion occurs in the Ames room shown in figure 7.31. In comparison to his 4-foot daughter, Tanya Zimbardo, your 6-foot-tall author looks quite short in the left corner of this room, but he looks enormous in the right corner. The reason for this illusion is that you perceive the room to be rectangular, with the two back corners equally distant from you. Thus you perceive Tanya's actual size as being consistent with the size of the images on your retina in both cases. In fact, Tanya is not at the same distance, because the Ames room creates a clever illusion. It appears to be a rectangular room, but it is actually made from nonrectangular surfaces at odd angles in depth and height, as you can see in the drawings that accompany the photos. Any person on the right will make a larger retinal image, because he or she is twice as close to the observer.

Another way that the perceptual system can infer objective size is by using prior knowledge about the characteristic size of similarly shaped objects. For instance, once you recognize the shape of a house, a tree, or a dog, you have a pretty good idea of how big each is, even without knowing its distance from you. Universal Studios in Hollywood uses your expectations about the normal sizes of doors to make its actors in westerns look bigger or smaller to you. The doors on one side of the street on a western set are made to be smaller than the

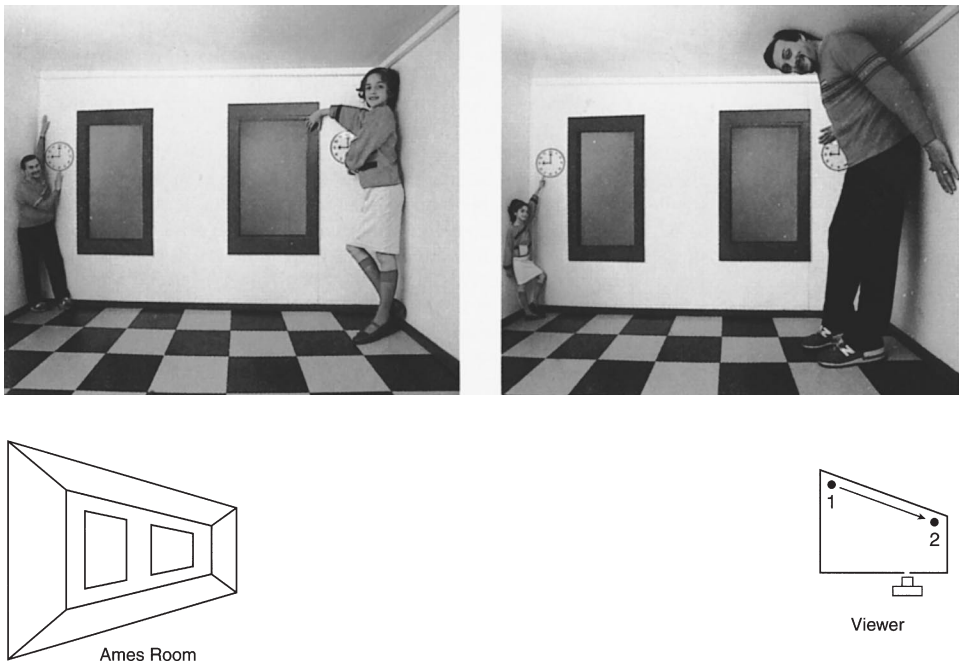


Figure 7.31
The Ames room.

doors on the other side of the street. When shooting the scenes of the westerns, directors position male actors on the side of the street with small doors. This makes them look bigger. Female actors, on the other hand, get filmed on the other side of the street, against the background of large doors, which makes them look petite.

When past experience does not give you knowledge of what familiar objects look like at extreme distances, size constancy may break down. You have experienced this problem if you have looked down at people from the top of a skyscraper and thought that they resembled ants. Consider, also, the experience of a man named Kenge of the equatorial Africa Pygmy culture. Kenge had lived in dense tropical forests all his life. He had occasion, one day, to travel by car for the first time across an open plain with anthropologist Colin Turnbull. Later, Turnbull described Kenge's reactions.

> Kenge looked over the plains and down to where a herd of about a hundred buffalo were grazing some miles away. He asked me what kind of *insects* they were, and I told him they were buffalo, twice as big as the forest buffalo known to him. He laughed loudly and told me not to tell such stupid stories, and asked me again what kind of insects they were. He then talked to himself, for want of more intelligent company, and tried to liken the buffalo to the various beetles and ants with which he was familiar.
>
> He was still doing this when we got into the car and drove down to where the animals were grazing. He watched them getting larger and

Figure 7.32
Shape constancy. As a coin is rotated, its image becomes an ellipse that grows narrower and narrower until it becomes a thin rectangle, an ellipsis again, and then a circle. At each orientation, however, it is still perceived as a circular coin.

> larger, and though he was as courageous as any Pygmy, he moved over and sat close to me and muttered that it was witchcraft.... Finally, when he realized that they were real buffalo he was no longer afraid, but what puzzled him still was why they had been so small, and whether they *really* had been small and had so suddenly grown larger, or whether it had been some kind of trickery. (Turnbull, 1961, p. 305)

In this unfamiliar perceptual environment, Kenge first tried to fit his novel perceptions into a familiar context, by assuming the tiny, distant specks he saw were insects. With no previous experience seeing buffalo at a distance, he had no basis for size constancy, and as the fast-moving car approached them and Kenge's retinal images got larger and larger, he had the frightening illusion that the animals were changing in size. We can assume that, over time, Kenge would have come to see them as Turnbull did. The knowledge he acquired would allow him to arrive at an appropriate perceptual interpretation for his sensory experience.

Shape constancy is closely related to size constancy. You perceive an object's actual shape correctly even when the object is slanted away from you, making the shape of the retinal image substantially different from that of the object itself. For instance, a rectangle tipped away projects a trapezoidal image onto your retina; a circle tipped away from you projects an elliptical image (see figure 7.32). Yet you usually perceive the shapes accurately as a circle and a rectangle slanted away in space. When there is good depth information available, your visual system can determine an object's true shape simply by taking into account your distance from its different parts.

Orientation Constancy When you tilted your head to the side in viewing your book, the world did not seem to tilt; only your own head did. *Orientation constancy* is your ability to recognize the true orientation of the figure in the real world, even though its orientation in the retinal image is changed. Orientation constancy relies on output from the vestibular system in your inner ear—which makes available information about the way in which your head is tilted. By combining the output of the vestibular system with retinal orientation, your visual system is usually able to give you an accurate perception of the orientation of an object in the environment.

In familiar environments, prior knowledge provides additional information about objective orientation. However, you may not be good at recognizing complex and unfamiliar figures when they are seen in unusual orientations.

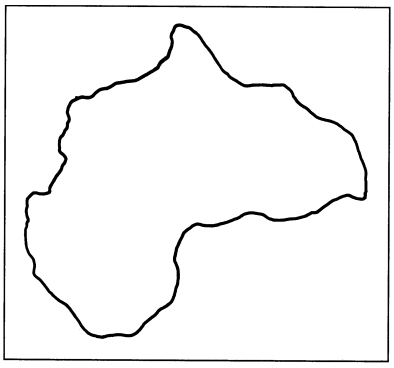


Figure 7.33
Africa rotated 90 degrees.

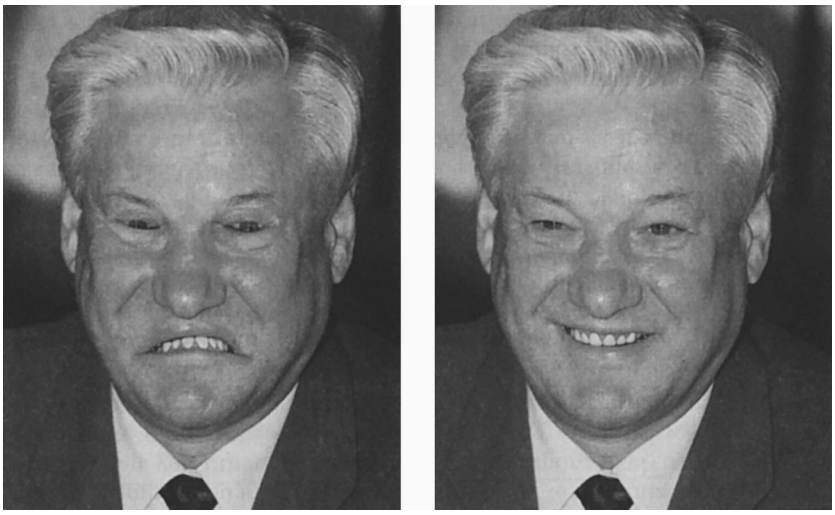


Figure 7.34
Which of these portraits might express Boris Yeltsin's feelings after hearing bad news about the Russian economy?

Can you recognize the shape in figure 7.33? When a figure is complex and consists of subparts, you must adjust for the orientation of each part separately (Rock, 1986). So, while you rotate one part to its proper orientation, other parts are still perceived as unrotated. Look at the two upside-down pictures of Russian leader Boris Yeltsin in figure 7.34. You can probably tell that one of them has been altered slightly around the eyes and mouth, but the two pictures look pretty similar. Now turn the book upside down and look again. The same pictures look extraordinarily different now. One is still Boris Yeltsin, but the other is a ghoulish monster that not even his mother could love! Your failure to see that obvious difference before turning the book upside down may be due to your inability to rotate all of the parts of the face at the same time. It is also a

function of years of perceptual training to see the world right side up and to perceive faces in their usual orientation.

Identification and Recognition Processes

You can think of all the perceptual processes described so far as providing reasonably accurate knowledge about physical properties of the distal stimulus—the position, size, shape, texture, and color of objects in a three-dimensional environment. With just this knowledge and some basic motor skills, you would be able to walk around without bumping into anything and manipulate objects that were small and light enough to move. However, you would not know what the objects were or whether you had seen them before. Your experience would resemble a visit to an alien planet where everything was new to you; you wouldn't know what to eat, what to put on your head, what to run away from, or what to date. Your environment appears nonalien because you are able to recognize and identify most objects as things you have seen before and as members of the meaningful categories that you know about from experience. Identification and recognition attach meaning to percepts.

Bottom-up and Top-down Processes

When you identify an object, you must match what you see against your stored knowledge. Taking sensory data into the system and sending it upward for extraction and analysis of relevant information is called bottom-up processing. *Bottom-up processing* is anchored in empirical reality and deals with bits of information and the transformation of concrete, physical features of stimuli into abstract representations. This type of processing is also called *data-driven* processing, because your starting point for identification is the sensory evidence you obtain from the environment—the data.

In many cases, however, you can use information you already have about the environment to help you make a perceptual identification. If you visit a zoo, for example, you might be a little more ready to recognize some types of animals than you otherwise would be. You are more likely to hypothesize that you are seeing a tiger than you would be in your own back yard. When your expectations affect perception, the phenomenon is called top-down processing. *Top-down processing* involves your past experiences, knowledge, motivations, and cultural background in perceiving the world. With top-down processing, higher mental functioning influences how you understand objects and events. Top-down processing is also known as *conceptually driven* (or hypothesis-driven) processing, because the concepts you have stored in memory are affecting your interpretation of the sensory data. The importance of top-down processing can be illustrated by drawings known as *droodles* (Price, 1953/1980). Without the labels, these drawings are meaningless. However, once the drawings are identified, you can easily find meaning in them (see figure 7.35).

For a more detailed example of top-down versus bottom-up processing, we turn to the domain of speech perception. You have undoubtedly had the experience of trying to carry on a conversation at a very loud party. Under those circumstances, it's probably true that not all of the physical signal you are producing arrives unambiguously at your acquaintance's ears: some of what you had to say was almost certainly obscured by coughs, thumping music, or peals

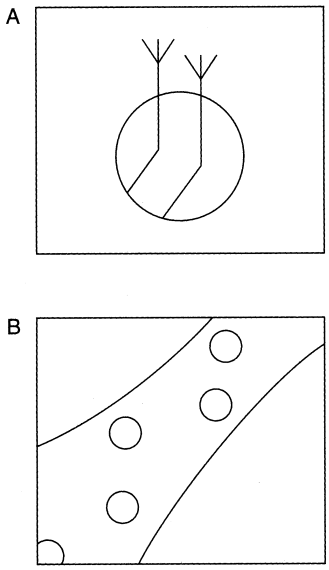


Figure 7.35
Droodles. What are these animals? Do you see in (A) an early bird who caught a very strong worm and in (B) a giraffe's neck? Each of these figures can be seen as representing something familiar to you, although this perceptual recognition usually does not occur until some identifying information is provided.

of laughter. Even so, people rarely realize that there are gaps in the physical signal they are experiencing. This phenomenon is known as *phonemic restoration* (Warren, 1970). Samuel (1981, 1991) has shown that subjects often find it difficult to tell whether they are hearing a word that has a noise replacing part of the original speech signal or whether they are hearing a word with a noise just superimposed on the intact signal (see the top panel of figure 7.36).

The bottom panel of figure 7.36 shows how bottom-up and top-down processes could interact to produce phonemic restoration (McClelland & Elman, 1986). Suppose part of what your friend says at a noisy party is obscured so that the signal that arrives at your ears is "I have to go home to walk my (noise)og." If noise covers the /d/, you are likely to think that you actually heard the full word *dog*. But why? In figure 7.36, you see two of the types of information relevant to speech perception. We have the individual sounds that make up words, and the words themselves. When the sounds /o/ and /g/ arrive in this system, they provide information—in a bottom-up fashion—to the word level (we have given only a subset of the words in English that end with /og/). This provides you with a range of candidates for what your friend might have said. Now top-down processes go to work—the context helps you select *dog* as the most likely word to appear in this utterance. When all of this happens swiftly enough—bottom-up identification of a set of candidate words and top-down selection of the likely correct candidate—you'll never know that the /d/ was missing. Your perceptual processes believe that the word was intact. (You may want to review figure 7.4 to see how everything in this chapter fits together.)

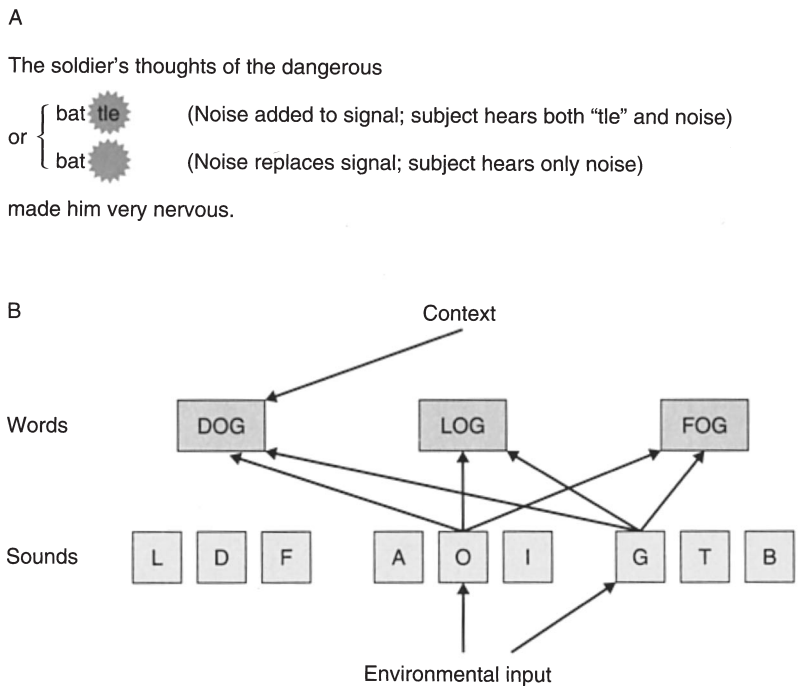


Figure 7.36
Phonemic restoration.

Object Recognition

From the example of speech perception, we can derive a general approach that researchers bring to the bottom-up study of recognition: they try to determine the building blocks that perceptual systems use to recognize whole percepts. For language, your speech perception processes combine environmental information about series of sounds to recognize individual words. What are the units from which you construct your representations of objects in the world? How, for example, do you decide that a gray, oddly shaped, medium-size, furry thing is actually a cat? Presumably, you have a memory representation of a cat. The identification process consists in matching the information in the percept to your memory representation of the cat. But how are these matches accomplished? One possibility is that the memory representations of various objects consist of components and information about the way these components are attached to each other (Marr & Nishihara, 1978). *Irving Biederman* (1985, 1987) has proposed that all objects can be assembled from a set of *geometrical ions*, or *geons*. Geons are not a large or arbitrary set of shapes. Biederman argued that a set of 36 geons can be defined by following the rule that each three-dimensional geon creates a unique pattern of stimulation on the two-dimensional retina. This uniqueness rule would allow you to work backward from a pattern of sensory stimulation to a strong guess at what the environmental object was like. Figure 7.37 gives examples of the way in which objects can be assembled from this collection of standard parts.

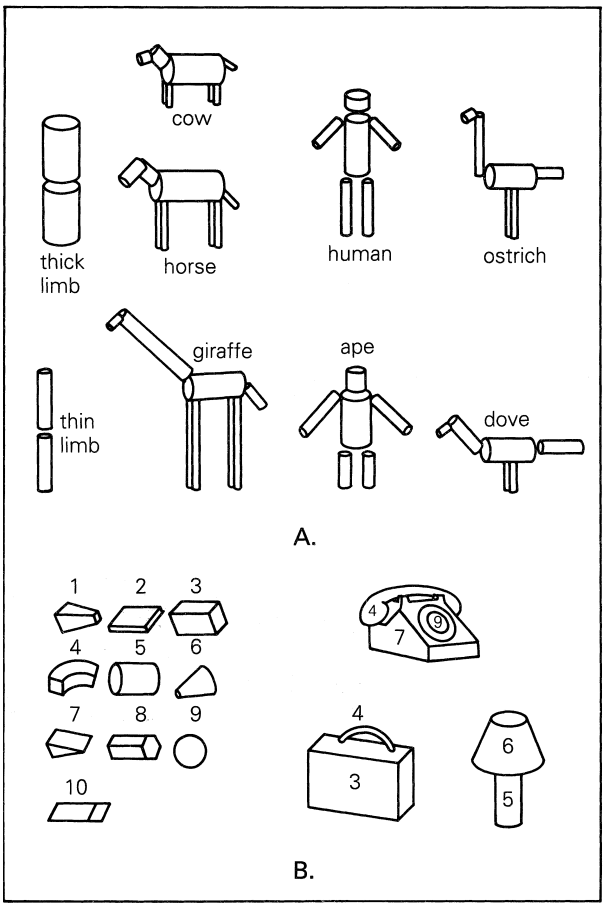


Figure 7.37
Recognition by components. Suggested components of 3-dimensional objects and examples of how they may combine. In the top half of the figure, each 3-D object is constructed of cylinders of different sizes. In the bottom half of the figure, several different building blocks are combined to form familiar objects.

Researchers have shown that such parts do, in fact, play a role in object recognition. They have done so by presenting subjects with degraded pictures of objects that either do or do not leave parts intact (Biederman, 1987; Biederman & Cooper, 1991). The first column of figure 7.38 shows line drawings of common objects. The middle column shows those same objects with only information deleted that still allows you to detect what the parts are and how they are combined. The right-hand column presents deletions that disrupt the identities of and relationships between the parts. Do you agree that it would be hard for you to recognize some of these objects based just on the drawings in the third column? The contrast here suggests that you can recognize objects with limited information (just as you can restore missing phonemes), but not if that information disrupts the critical parts.

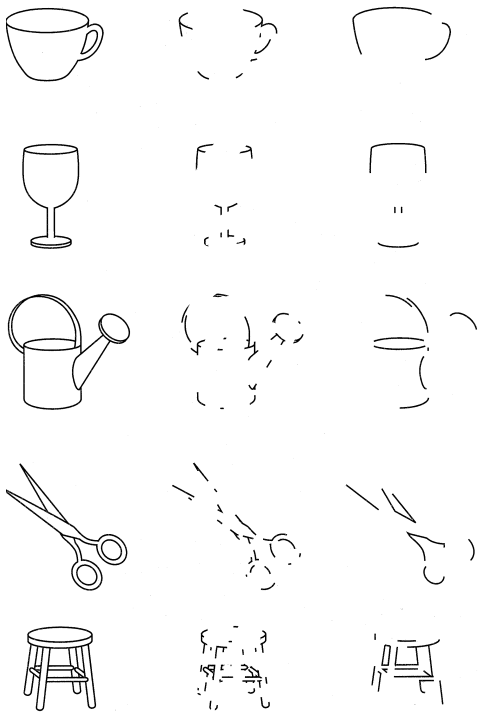


Figure 7.38
Role of parts in object recognition. The deletions of visual information in the middle column leave the parts intact. In the right-hand column, the deletions disrupt the parts. Do you agree that the objects are easier to recognize in the middle versions?

Recovery of components alone, however, will not always be sufficient to recognize an object (Tarr, 1994). One difficulty, as shown in figure 7.39, is that you often see objects from radically different perspectives. The appearance of the parts that make up the object may be quite different from each of these perspectives. As a hedge against this difficulty, you must store separate memory representations for each of the major perspectives from which you view standard objects (Tarr & Pinker, 1989). When you encounter an object in the environment, you may have to mentally transform the percept to determine if it correctly matches one of those views. Thus to recognize a gray, oddly shaped, medium-size, furry thing as a cat, you must recognize it both as an appropriate combination of geons and as that appropriate combination of geons from a specific viewpoint.

The Influence of Contexts and Expectations

What also might help you recognize the cat, however, is to find that gray, oddly shaped, medium-size, furry thing in its accustomed place in your home. This is the top-down aspect of perception: expectations can influence your hypotheses about what is out there in the world. Have you ever had the experience of seeing people you knew in places where you didn't expect

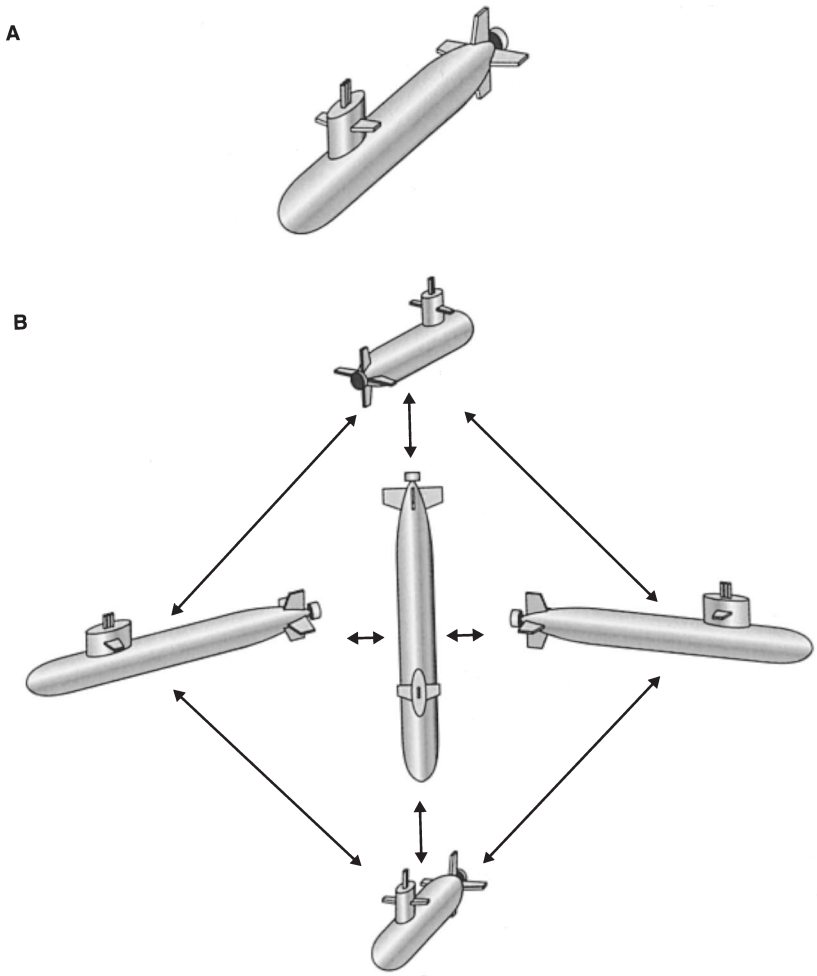


Figure 7.39
Looking at the same object from different positions. You see different parts of an object when you view it from different perspectives. To overcome this difficulty, you store multiple views of complex objects in memory.

to see them, such as in the wrong city or the wrong social group? It takes much longer to recognize them in such situations, and sometimes you aren't even sure that you really know them. The problem is not that they look any different but that the *context* is wrong; you didn't *expect* them to be there. The spatial and temporal context in which objects are recognized provides an important source of information, because from the context you generate expectations about what objects you are and are not likely to see nearby (Biederman, 1989).

Perceptual identification depends on your expectations as well as on the physical properties of the objects you see—*object identification is a constructive, interpretive process.* Depending on what you already know, where you are, and

what else you see around you, your identification may vary. Read the following words:

THE CAT

They say *THE CAT*, right? Now look again at the middle letter of each word. Physically, these two letters are exactly the same, yet you perceived the first as an H and the second as an A. Why? Clearly, your perception was affected by what you know about words in English. The context provided by T_E makes an H highly likely and an A unlikely, whereas the reverse is true of the context of C_T (Selfridge, 1955).

Researchers have often documented the effects of context and expectation on your perception (and response) by studying set. *Set* is a temporary readiness to perceive or react to a stimulus in a particular way. There are three types of set: motor, mental, and perceptual. A *motor set* is a readiness to make a quick, prepared response. A runner trains by perfecting a motor set to come out of the blocks as fast as possible at the sound of the starting gun. A *mental set* is a readiness to deal with a situation, such as a problem-solving task or a game, in a way determined by learned rules, instructions, expectations, or habitual tendencies. A mental set can actually prevent you from solving a problem when the old rules don't seem to fit the new situation. A *perceptual set* is a readiness to detect a particular stimulus in a given context. A new mother, for example, is perceptually set to hear the cries of her child.

Often a set leads you to change your interpretation of an ambiguous stimulus. Consider these two series of words:

FOX; OWL; SNAKE; TURKEY; SWAN; D?CK

BOB; RAY; DAVE; BILL; HENRY; D?CK

Did you read through the lists? What word came to mind for D? CK in each case? If you thought DUCK and DICK, it's because the list of words created a set that directed your search of memory in a particular way.

Labels can provide a context that gives a perceptual set for an ambiguous figure. You have seen how meaningless droodles turn into meaningful objects. Look carefully at the picture of the woman in figure 7.40A; have a friend (but not you) examine figure 7.40B. Next, together look at figure 7.40C—what does each of you see? Did the prior exposure to the unambiguous pictures with their labels have any effect on perception of the ambiguous image? This demonstration shows how easy it is for people to develop different views of the same person or object, based on prior conditions that create different sets.

All the effects of context on perception clearly require that your memory be organized in such a fashion that information relevant to particular situations becomes available at the right times. In other words, to generate appropriate (or inappropriate) expectations, you must be able to make use of prior knowledge stored in memory. Sometimes you "see" with your memory as much as you see with your eyes.

Figure 7.40A
A young beauty.

Figure 7.40B
An old woman.

Figure 7.40C
Now what do you see?

Creatively Playful Perception

Because of your ability to go beyond the sensory gifts that evolution has bestowed on the human species, you can become more creative in the way you perceive the world. Your role model is not a perfectly programmed computerized robot with exceptional sensory acuity. Instead, it is a great artist like Pablo Picasso. Picasso's genius was, in part, attributable to his enormous talent for "playful perception." This artist could free himself from the bonds of perceptual and mental sets to see not the old in the new but the new in the old, the novel in the familiar, and the unusual figure concealed within the familiar ground.

Perceptual creativity involves experiencing the world in ways that are imaginative, personally enriching, and fun (Leff, 1984). You can accomplish perceptual creativity by consciously directing your attention and full awareness to the objects and activities around you. Your goal should be to become more flexible in what you allow yourself to perceive and think, remaining open to alternative responses to situations.

We can think of no better way to conclude this formal presentation of the psychology of perception than by proposing ten suggestions for playfully enhancing your powers of perception:

- Imagine that everyone you meet is really a machine designed to look humanoid, and all machines are really people designed to look inanimate.
- Notice all wholes as ready to come apart into separately functioning pieces that can make it on their own.
- Imagine that your mental clock is hooked up to a video recorder that can rewind, fast-forward, and freeze time.
- Recognize that most objects around you have a "family resemblance" to other objects.
- View the world as if you were an animal or a home appliance.
- Consider one new use for each object you view (use a tennis racket to drain cooked spaghetti).
- Suspend the law of causality so that events just happen, while coincidence and chance rule over causes and effects.
- Dream up alternative meanings for the objects and events in your life.
- Discover something really interesting about activities and people you used to find boring.
- Violate some of the assumptions that you and others have about what you would and wouldn't do (without engaging in a dangerous activity).

Final Lessons

The important lesson to be learned from the study of perception is that a perceptual experience in response to a stimulus event is a response of the whole person. In addition to the information provided when your sensory receptors are stimulated, your final perception depends on who you are, whom you are with, and what you expect, want, and value. A perceiver often plays two different roles that we can compare to gambling and interior design. As a gambler, a perceiver is willing to bet that the present input can be understood in terms of past knowledge and personal theories. As a compulsive interior deco-

rator, a perceiver is constantly rearranging the stimuli so that they fit better and are more coherent. Incongruity and messy perceptions are rejected in favor of those with clear, clean, consistent lines.

If perceiving were completely bottom-up, you would be bound to the same mundane, concrete reality of the here and now. You could register experience but not profit from it on later occasions, nor would you see the world differently under different circumstances. If perceptual processing were completely top-down, however, you could become lost in your own fantasy world of what you expect and hope to perceive. A proper balance between the two extremes achieves the basic goal of perception: to experience what is out there in a way that maximally serves your needs as a biological and social being moving about and adapting to your physical and social environment.

Recapping Main Points

Sensing, Organizing, Identifying, and Recognizing

Your perceptual systems do not simply record information about the external world but actively organize and interpret information as well. Perception is a three-stage process consisting of a sensory stage, a perceptual organization stage, and an identification and recognition stage. At the sensory level of processing, physical energy is detected and transformed into neural energy and sensory experience. At the organizational level, brain processes organize sensations into coherent images and give you perception of objects and patterns. At the level of identification, percepts of objects are compared with memory representations in order to be recognized as familiar and meaningful objects. The task of perception is to determine what the distal (external) stimulus is from the information contained in the proximal (sensory) stimulus. Ambiguity may arise when the same sensory information can be organized into different percepts. Knowledge about perceptual illusions can give you clues about normal organizing processes.

Attentional Processes

Attention refers to your ability to select part of the sensory input and disregard the rest. Both your personal goals and the properties of objects in the world determine where you will focus your attention. Attention accomplishes its tasks by both facilitating the processing of the relevant, attended stimuli and suppressing the processing of irrelevant, unattended stimuli. Preattentive processing enables you to search the visual environment efficiently, although focused attention is required in many cases to find combinations of features. Attention also allows simple physical properties of objects to be combined correctly.

Organizational Processes in Perception

Organizational processes provide percepts consistent with the sensory data. These processes segregate your percepts into regions and organize them into figures that stand out against the ground. You tend to see incomplete figures as wholes; group items by similarity; and see "good" figures more readily. You tend to organize and interpret parts in relation to the spatial and temporal

context in which you experience them. You also tend to see a reference frame as stationary and the parts within it as moving, regardless of the actual sensory stimulus. In converting the two-dimensional information on the retina to a perception of three-dimensional space, the visual system gauges object size and distance: distance is interpreted on the basis of known size, and size is interpreted on the basis of various distance cues. You tend to perceive objects as having stable size, shape, and orientation. Prior knowledge normally reinforces these and other constancies in perception; under extreme conditions, perceptual constancy may break down.

Identification and Recognition Processes

During the final stage of perceptual processing—identification and recognition of objects—percepts are given meaning through processes that combine bottom-up and top-down influences. Context, expectations, and perceptual sets may guide recognition of incomplete or ambiguous data in one direction rather than another, equally possible one. Perception thus depends on what you know and expect as well as on the sensory stimulus.

References

Beck, J. (1972). Similarity groupings and peripheral discriminality under uncertainty. *American Journal of Psychology, 85*, 1–20.

Beck, J. (Ed.) (1982). *Organization and representation in perception.* Hillsdale, NJ: Erlbaum.

Biederman, I. (1985). Recognition by components: A theory of object recognition. *Computer Vision Graphics and Image Processing, 32*, 29–73.

Biederman, I. (1987). Recognition by components. *Psychological Review, 94*, 115–147.

Biederman, I. (1989). Higher-level vision. In D. N. Osherson, H. Sasnik, S. Kosslyn, K. Hollerbach, E. Smith, & N. Block (Eds.), *An invitation to cognitive science*. Cambridge, MA: MIT Press.

Biederman, I., & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23*, 393–419.

Bregman, A. S. (1982). Asking the "what for" question in auditory perception. In M. Kobovy & J. Pomerantz (Eds.), *Perceptual organization* (pp. 99–118). Hillsdale, NJ: Erlbaum.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 25*, 975–979.

Cohen, S., & Girgus, J. S. (1973). Visual spatial illustrations: Many explanations. *Science, 179*, 503–504.

Cutting, J., & Proffitt, D. (1982). The minimum principle and the perception of absolute, common, and relative motions. *Cognitive Psychology, 14*, 211–246.

Driver, J., & Tipper, S. (1989). On the nonselectivity of "selective" seeing: Contrasts between interference and priming in selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 304–314.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gibson, J. J. (1979). *An ecological approach to visual perception*. Boston: Houghton Mifflin.

Hillstrom, A. P., & Yantis, S. (1994). Visual motor and attentional capture. *Perception & Psychophysics, 55*, 399–411.

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23*, 420–456.

Julesz, B. (1981a). Figure and ground perception in briefly presented isodipole textures. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 27–54). Hillsdale, NJ: Erlbaum.

Julesz, B. (1981b). Textons, the elements of texture perception and their interaction. *Nature, 290*, 91–97.

Kanizsa, G. (1979). *Organization in vision*. New York: Praeger.

Leff, H. (1984). *Playful perception: Choosing how to experience your world*. Burlington, VT: Waterfront Books.

Mace, W. M. (1977). James J. Gibson's strategy for perceiving. Ask not what's inside your head, but what your head's inside of. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing.* Hillsdale, NJ: Erlbaum.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Marr, D., & Nishihara, H. K. (1978). Representations and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London (Series B), 200*, 269–294.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–18.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*, 164–488.

Palmer, S. (1989). Reference frames in the perception of shape and orientation. In B. Shepp & M. Ballisteros (Eds.), *Object perception* (pp. 121–163). Hillsdale, NJ: Erlbaum.

Palmer, S. E. (1984). The psychology of perceptual organization: A transformational approach. In A. Rosenfeld & J. Beck (Eds.), *Human and machine vision*. New York: Academic Press.

Pittenger, J. B. (1988). Direct perception of change. *Perception, 17*, 119–133.

Pomerantz, J., & Kubovy, M. (1986). Theoretical approaches to perceptual oganization. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 3, pp. 1–46). New York: Wiley.

Price, R. (1953/1980). *Droodles*. Los Angeles: Price/Stern/Sloan.

Rock, I. (1983). *The logic of perception*. Cambridge, MA: Bradford Books/MIT Press.

Rock, I. (1986). The description and analysis of object and event perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 33–71). New York: Wiley.

Rock, I., & Gutman, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 275–285.

Selfridge, O. G. (1955). Pattern recognition and modern computers. *Proceedings of the Western Joint Computer Conference*. New York: Institute of Electrical and Electronics Engineers.

Shaw, R., & Turvey, M. T. (1981). Coalitions as models for ecosystems: A realist perspective on perceptual organization. In M. Kubovey & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 343–346). Hillsdale, NJ: Erlbaum.

Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imaging, thinking and dreaming. *Psychological Review, 91*, 417–447.

Shepard, R. N., & Johnson, D. S. (1984). Auditory illusions demonstrating that tomes are assimilated to an internalized musical scale. *Science, 226*, 1333–1334.

Shepp, B., & Ballisteros, M. (Eds.) (1989). *Object perception*. Hillsdale, NJ: Erlbaum.

Tarr, M. J. (1994). Visual representation: From features to objects. In V. S. Ramachandran (Ed.), *The encyclopedia of human behavior*. San Diego: Academic Press.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21*, 233–282.

Tipper, S. P., & Driver, J. (1988). Negative priming between pictures and words in a selective attention task: Evidence for semantic processing of ignored stimuli. *Memory and Cognition, 16*, 64–70.

Tipper, S. P., Weaver, B., Cameron, S., Brehaut, J. C., & Bastedo, J. (1991). Inhibitory mechanisms of attentions in identification and localization tasks: Time course and disruption. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 681–692.

Todrank, J., & Bartoshuk, L. M. (1991). A taste illusion: Taste sensation localized by touch. *Physiology and Behavior, 50*, 1027–1031.

Treisman, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12*, 242–248.

Treisman, A. (1986). Properties, parts and objects. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2). New York: Wiley.

Treisman, A. (1988). Features and objects. The fourteenth Bartlett Memorial Lecture. *The Quarterly Journal of Experimental Psychology, 40*, 20–237.

Treisman, A. (1992). Perceiving and re-perceiving objects. *American Psychologist, 47*, 862–875.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 459–478.

Turnbull, C. (1961). *The forest people*. New York: Simon and Schuster.

Vasari, G. (1967). *Lives of the most eminent painters*. New York: Heritage.

Wertheimer, M. (1923). Untersuchengen zur lehre von der gestalt, II. *Psychologische Forschung, 4*, 301–350.

Wolfe, J. M. (1992). The parallel guidance of visual attention. *Current Directions in Psychological Science, 1*, 124–128.

Wolfe, J. M., Friedman-Hill, S. R., & Bilsky, A. B. (1994). Parallel processing of part-whole information in visual search tasks. *Perception & Psychophysics, 55*, 537–550.

Yantis, S. (1993). Stimulus-driven attentional capture. *Current Directions in Psychological Science, 2*, 156–161.

Chapter 8

Organizing Objects and Scenes

Stephen E. Palmer

The Problem of Perceptual Organization The concept of perceptual organization originated with Gestalt psychologists early in this century. It was one of the central concepts in their attack on the atomistic assumption of Structuralism. The Structuralists conceived of visual perception as a simple concatenation of sensory "atoms" consisting of pointlike color sensations. This view of visual perception is extremely local in the sense that each atom was defined by a particular retinal position and thought to be independent of all other atoms, at least until they were bound together into larger spatial complexes by the process of associative learning. The Gestaltists, in contrast, believed that visual perception arose from global interactions within the visual nervous system and resulted from the overall structure of visual stimulation itself. "Perceptual organization" was the name they used to refer both to this theoretical idea and to the set of phenomena they discovered in support of it.

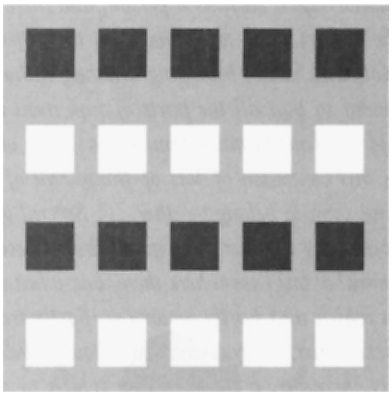
Max Wertheimer, one of the founding fathers of Gestalt psychology, first posed the problem of perceptual organization. He asked how people are able to perceive a coherent visual world that is organized into meaningful objects rather than the chaotic juxtaposition of different colors that stimulate the individual retinal receptors. His point can perhaps be most easily understood by considering what the output of the retinal mosaic would be for a simple but highly structured image. Figure 8.1A illustrates such an output as a numerical array, in which each number represents the neural response of a single retinal receptor. In this numerical form, it is nearly impossible to grasp the structure and organization of the image without extensive scrutiny. This situation is a lot like the one the visual system faces in trying to organize visual input, because the structure we perceive so effortlessly is not explicitly given in the stimulus image but must be discovered by the visual nervous system. In fact, there is a potentially limitless number of possible organizations in an image, only one of which we typically perceive. Which one we experience and why we perceive it rather than others are thus questions that require explanations.

The structure of the numerical image becomes completely obvious when you see these same values as luminance levels, as illustrated in figure 8.1B. It is a picture of several black and white squares that are organized into four horizontal rows on a gray background. But why is this simple structure so obvious when we view the image and so obscure when we look at the array of num-

From chapter 6 in *Vision Science: Photons to Phenomenology* (Cambridge, MA: MIT Press, 1999), 255–269. Reprinted with permission.

5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	2	2	5	2	2	5	2	2	5	2	2	5	2	2	5
5	2	2	5	2	2	5	2	2	5	2	2	5	2	2	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	8	8	5	8	8	5	8	8	5	8	8	5	8	8	5
5	8	8	5	8	8	5	8	8	5	8	8	5	8	8	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	2	2	5	2	2	5	2	2	5	2	2	5	2	2	5
5	2	2	5	2	2	5	2	2	5	2	2	5	2	2	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	8	8	5	8	8	5	8	8	5	8	8	5	8	8	5
5	8	8	5	8	8	5	8	8	5	8	8	5	8	8	5
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

A



B

Figure 8.1
The problem of perceptual organization. When an optical image is registered on the retina, the visual system is faced with trying to find structure in the pattern of receptor outputs, depicted in part A by a numerical array in which high numbers correspond to light regions and low numbers to dark regions. When observers view the corresponding gray-scale image (B), they immediately and effortlessly organize it into four rows of light and dark squares against a gray background.

bers? The reason is that the human visual system has evolved to learn how to detect edges, regions, objects, groups, and patterns from the structure of luminance and color in optical images. The gray-scale image in figure 8.1B engages these mechanisms fully, whereas the numerical image scarcely does at all. The same information is present in both images, of course, but the numerical image comes in a form that the visual system cannot discern directly. A theorist who is trying to explain visual perception is in much the same position as you are in trying to find structure in the numerical image: None of the organization that the visual system picks up so automatically and effortlessly can be presupposed, since that is the very structure that must be explained.

Why does visual experience have the organization it does? The most obvious answer is that it simply reflects the structure of the external world. By this ac-

count, the physical environment actually consists of things like surfaces and objects arranged in space rather than points of color, and this is why perception is organized as it is. This is the naive realist's answer, and there is undoubtedly something to it. Surely evolutionary utility requires that perceptual organization reflect structure in the organism's environment, or at least the part of it that is relevant to the organism's survival. Imagine, for example, how much less useful vision would be if it characteristically misorganized the world. But although the naive realist's answer might help explain perceptual organization in an evolutionary sense—*why* perceptual experience has the structure it does—it does not explain the mechanisms of organization: *how* it unfolds in time during acts of perception. The goal of this chapter is to shed light on these mechanisms and the stimulus factors that engage them.

The Experience Error The major difficulty with the view of naive realism is that the visual system does not have direct access to facts about the environment; it has access only to facts about the image projected onto the retina. That is, an organism cannot be presumed to know how the environment is structured except through sensory information. The Gestaltists referred to the naive realist's approach to the problem of perceptual organization as the *experience error* because it arises from the false (and usually implicit) assumption that the structure of perceptual experience is somehow directly given in the array of light that falls on the retinal mosaic (Köhler, 1947). This optic array actually contains an infinite variety of possible organizations, however, only one of which the visual system usually achieves.

The confusion that underlies the experience error is typically to suppose that the starting point for vision is the distal stimulus rather than the proximal stimulus. This is an easy trap to fall into, since the distal stimulus is an essential component in the causal chain of events that normally produces visual experiences. It also corresponds to the interpretation the visual system strives to achieve. Taking the distal stimulus as the starting point for vision, however, seriously underestimates the difficulty of visual perception because it presupposes that certain useful and important information comes "for free." But the structure of the environment is more accurately regarded as the *result* of visual perception rather than its starting point. As obvious and fundamental as this point might seem, now that we are acquainted with the difficulties in trying to make computers that can "see," the magnitude of the problem of perceptual organization was not fully understood until Wertheimer raised it in his seminal paper in 1923. Indeed, although significant progress has been made in the intervening years, vision scientists are still uncovering new layers of this important and pervasive problem.

8.1 Perceptual Grouping

Wertheimer's initial assault on the problem of perceptual organization was to study the stimulus factors that affect *perceptual grouping*: how the various elements in a complex display are perceived as "going together" in one's perceptual experience. He approached this problem by constructing very simple arrays of geometric elements and then varying the stimulus relations among

them to determine which ones caused certain elements to be grouped together perceptually.

Logically, a set of elements can be partitioned in a number of different ways, corresponding to the number of possible ways of dividing them into mutually exclusive subsets. This number becomes very large very quickly: For 10 elements, there are 42 possible groupings; but for 100 elements, there are 190,569,292. The number of logically possible groupings is even larger than the number of partitions if one considers hierarchical embedding of subsets and/or overlap among their members. Psychologically, however, only one of these groupings is perceived at one time, and the first one is usually the only one. How does this happen? And what properties of the stimulus image determine which grouping people perceive?

8.1.1 The Classical Principles of Grouping

In his investigations, Wertheimer started with a single line of equally spaced dots as shown in figure 8.2A. These dots do not group together into any larger perceptual units—except the line of dots as a whole. He then noted that when he altered the spacing between adjacent dots so that some pairs were closer together and others were farther apart, as in figure 8.2B, the closer ones grouped

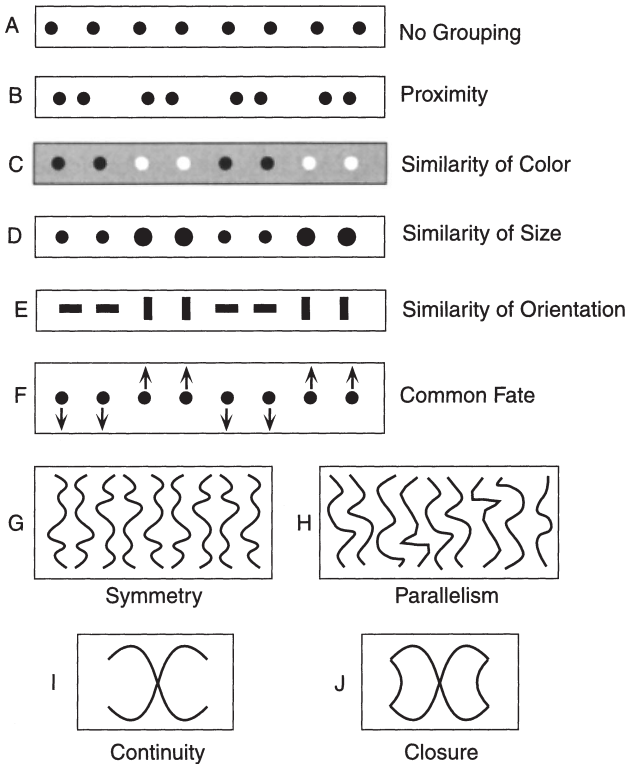


Figure 8.2
Classical principles of grouping. Gestalt psychologists identified many different factors that govern which visual elements are perceived as going together in larger groups. (See text for details.)

strongly together into pairs. This factor of relative closeness, which Wertheimer called *proximity*, was the first of his famous *laws of grouping*. (From now on, we will refer to them as "principles" or "factors" of grouping because, as we will see, they are considerably weaker than one would expect of scientific laws.) The evidence that he offered for the potency of proximity as a factor in grouping was purely phenomenological. He simply presented the array in figure 8.2B to his readers and appealed directly to their experiences of which dots they saw as "going together." Since nobody has ever seriously disputed Wertheimer's claim that the closer dots group perceptually, the principle of proximity was thereby firmly established simply by demonstration, without any formal experiment.

It is perhaps worth making a brief digression here concerning the phenomenological methods employed by Gestalt psychologists. Their demonstrations have often been criticized because they lack the rigorous experimental procedures adhered to by behaviorally oriented researchers (e.g., Pomerantz & Kubovy, 1986). In actuality, however, the Gestaltists were often able to bypass formal experiments simply because the phenomena that they discovered were so powerful that no experiment was needed. If hundreds or even thousands of people viewing their displays agree with their claims about the resulting phenomenological impression, why bother with a formal experiment? As Irvin Rock often remarked, the demonstrations of Gestalt psychologists, such as those in figure 8.2, can actually be viewed as ongoing experiments with an indefinitely large number of subjects—of which you are now one—virtually all of whom "show the effect." In cases for which the facts were less clear, Gestalt psychologists often performed perfectly reasonable experiments and recorded objective data, such as the number of observers who reported one percept versus another (e.g., Goldmeier, 1936/1972). Thus, their phenomenological methods are not as far removed from modern behavioral ones as is often suggested.

After demonstrating the effect of proximity, Wertheimer went on to illustrate many of the other principles of grouping portrayed in figure 8.2. Figures 8.2C, 8.2D, and 8.2E, for example, demonstrate the principle of *similarity*: All else being equal, the most similar elements (in color, size, and orientation in these examples) tend to be grouped together. Similarity can thus be considered a very general principle of grouping because it covers many different properties.

Another powerful factor is what Wertheimer called *common fate*: All else being equal, elements that move in the same way tend to be grouped together. Although this cannot be demonstrated in a static display, grouping by common fate is indicated symbolically by the arrows in figure 8.2F. Notice that common fate can actually be considered a special case of similarity grouping in which the similar property is velocity of movement. It has even been claimed that proximity can be considered a special case of similarity grouping in which the underlying dimension of similarity is the position of the elements.

Not all possible similarities are equally effective, however, and some do not produce much grouping at all. Consider the row of V's in figure 8.3A, for example. Adjacent pairs differ by 180° in orientation, yet there is very little spontaneous grouping by similarity in this display. Figure 8.3B shows the same figures in pairs that differ by only 45° in orientation, and now the pairwise grouping is immediately apparent. The visual system thus seems to be much more sensitive to certain kinds of differences than to others. Even subtle differ-

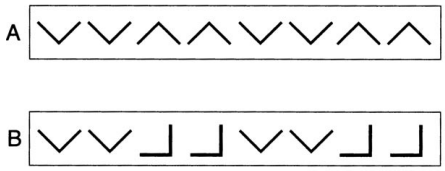


Figure 8.3
Degrees of grouping. Not all factors are equally effective in producing grouping. In part A, elements that differ by 180° in orientation are not strongly grouped, whereas those in part B that differ by only 45° produce strong grouping.

ences like those in figure 8.3A can be perceived by deliberate scrutiny involving focused attention, but such processes appear to be different from normal effortless grouping such as occurs in viewing figure 8.3B.

Gestalt psychologists also described several further factors that influence perceptual grouping of linelike elements. Symmetry (figure 8.2G) and parallelism (figure 8.2H), for example, are factors that influence the grouping of individual lines and curves. Figure 8.2I illustrates the important factor of *good continuation* (or *continuity*) of lines or edges: All else being equal, elements that can be seen as smooth continuations of each other tend to be grouped together. Its effect is manifest in this figure because observers perceive it as containing two continuous intersecting lines rather than as two angles whose vertices meet at a point. Figure 8.2J illustrates the further factor of *closure*: All else being equal, elements forming a closed figure tend to be grouped together. Note that this display shows that closure can overcome continuity because the very same lines that were organized as two intersecting lines in part I are organized as two angles meeting at a point in part J. According to Wertheimer's analysis, this is because the noncontinuous segments now constitute parts of the same closed figure.

The demonstrations of continuity and closedness in figures 8.2I and 8.2J illustrate an important limitation in current knowledge about grouping principles. As formulated by Gestalt psychologists, they are *ceteris paribus rules*, which means that they can predict the outcome of grouping with certainty only *when everything else is equal*—that is, when there is no other grouping factor influencing the outcome. We saw, for example, that continuity governs grouping when the elements do not form a closed figure, but it can be overcome by closure when they do.

The difficulty with ceteris paribus rules is that they provide no general purpose scheme for integrating several potentially conflicting factors into an overall outcome—that is, for predicting the strength of their combined influences. The same problem arises for all the previously mentioned principles of grouping. If proximity influences grouping toward one outcome and similarity in color toward another, the grouping that will be perceived depends heavily on the particular example. Figure 8.4A shows a case in which proximity is strong enough to overcome color similarity, whereas figure 8.4B shows one in which color similarity dominates. The visual system clearly integrates over many grouping factors, but we do not yet understand how it does so. Later in this

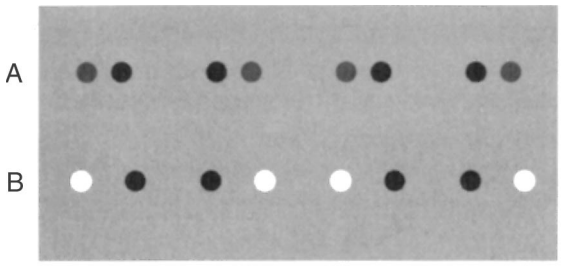


Figure 8.4
Tradeoffs between grouping by color and proximity. Large differences in proximity and small differences in color lead to grouping by proximity, whereas large differences in color and small differences in proximity lead to grouping by color.

chapter we will describe a recent theory that is able to integrate several different aspects of similarity grouping in the process of texture segregation, but it cannot yet handle other grouping principles such as common fate, continuity, and closure.

8.1.2 New Principles of Grouping

There has been surprisingly little modern work on principles of perceptual grouping in vision. Recently, however, three new grouping factors have been proposed: *synchrony* (Palmer & Levitin, submitted), *common region* (Palmer, 1992) and *element connectedness* (Palmer & Rock, 1994a).

The principle of *synchrony* states that, all else being equal, visual events that occur at the same time will tend to be perceived as going together. Although this factor has previously been acknowledged as important in auditory perception (e.g., Bregman, 1978), it has not been systematically studied in vision until recently (Palmer & Levitin, submitted). Figure 8.5 depicts an example. Each element in an equally spaced row of black and white dots flickers at a given rate between black and white. The arrows indicate that half the circles change from black to white or from white to black at one time and the other half at a different time. When the alternation rate is about 25 changes per second or less, observers see the dots as strongly grouped into pairs based on synchrony. At faster rates, there is no grouping in what appears to be chaotic flickering of the dots. At very slow rates there is momentary grouping into pairs at the moment of change, but it dissipates during the constant interval between flickers. Synchrony is related to the classical principle of common fate in the sense that it is a dynamic factor, but as this example shows, the "fate" of the elements does not have to be common—some dots get brighter, and others get dimmer—as long as the change occurs at the same time.

Another recently identified principle of grouping is common region (Palmer, 1992). *Common region* refers to the fact that, all else being equal (ceteris paribus), elements that are located within the same closed region of space will be grouped together. Figure 8.6A shows an example that is analogous to Wertheimer's classic demonstrations (figures 8.2B–8.2E): A line of otherwise equivalent, equally spaced dots is strongly organized into pairs when they are enclosed within the

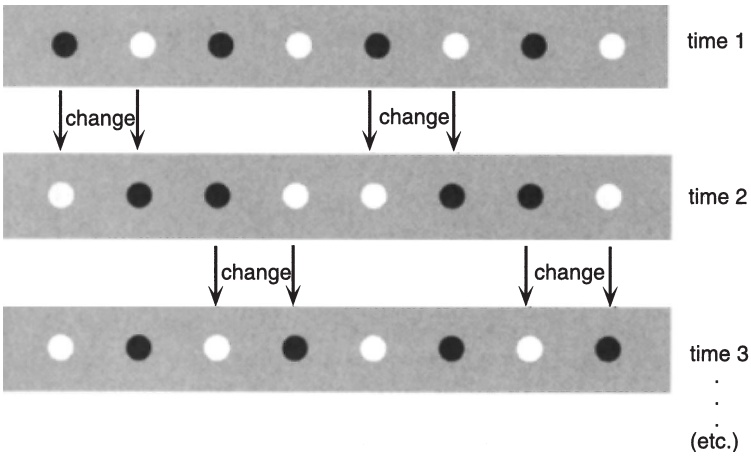


Figure 8.5
Grouping by synchrony. All else being equal, elements that change their properties at the same time (as indicated by the arrows) are grouped together.

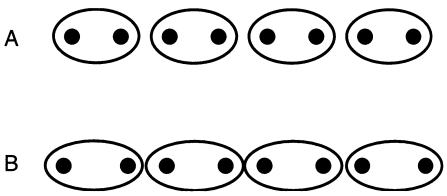


Figure 8.6
Grouping by common region. All else being equal, elements within the same region of space are grouped together (A), even when they are farther apart than elements in different regions (B). (After Palmer, 1992.)

same surrounding contour. Figure 8.6B shows that grouping by common region is powerful enough to overcome proximity that would, in itself, produce the opposite grouping structure.

A third newly proposed principle of grouping is *element connectedness*: All else being equal, elements that are connected by other elements tend to be grouped together. Palmer and Rock (1994) provide a number of demonstrations of its potency in grouping. An example that is analogous to Wertheimer's classic demonstrations is shown in figure 8.7A. The line of equally spaced dots is strongly grouped when subsets of the dots are connected by additional elements, such as the short horizontal line segments of this example. Figure 8.7B demonstrates that element connectedness can overcome even the powerful effect of proximity.

Wertheimer may not have considered element connectedness as a separate principle because it could be considered as the limiting case of maximal proximity. However, Palmer and Rock argue for distinguishing connectedness from proximity for several reasons. First, there is an important qualitative distinction between actual connectedness and mere proximity. Indeed, this distinction is a

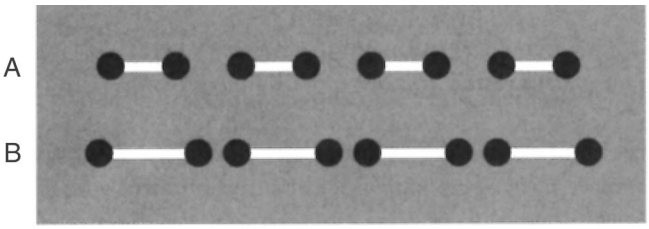


Figure 8.7
Grouping by element connectedness. All else being equal, elements that are connected to each other via additional elements are grouped together (A), even when they are farther apart than elements in different regions (B). (After Palmer & Rock, 1994a.)

cornerstone of the mathematical field of topology. Second, they note that what "goes together" in the strongest physical sense are those pieces of matter that are actually connected, not those that are merely close together. Parts of objects that are connected are much more closely coupled in their physical behavior than are two nearby objects, no matter how close they may be. Therefore, it makes sense for the visual system to be especially sensitive to connectedness as an indication of how to predict what will happen in the world. Third, there is an important phenomenological difference between connected and merely nearby objects. Element connectedness usually results in the perception of a single, unified object consisting of different parts, whereas mere proximity results in the perception of a looser aggregation of several separate but related objects. For these reasons, Palmer and Rock argued that proximity should be viewed as derivative from connectedness rather than the other way around.

The difference between the effects of mere proximity and those of actual connectedness suggests that the principles of grouping may not be a homogeneous set. In some cases, they result in *element aggregations*: loose confederations of objects that result from perceptual grouping operations. Proximity, similarity, common region, and certain cases of common fate often produce element aggregations in which the elements retain a high degree of perceptual independence despite their interrelation within the group. Other principles of grouping can produce *unit formation*: perception of a single, perceptually connected object from multiple underlying elements. Element connectedness, good continuation, and other cases of common fate frequently produce this more coherent organization into single unified objects.

One might think from the discussion of grouping principles that they are mere textbook curiosities, only distantly related to anything that occurs in normal perception. Wertheimer claimed, however, that they pervade virtually all perceptual experience because they are responsible for determining the objects and parts we perceive in the environment. Some dramatic examples of where perceptual organization goes wrong can be identified in natural camouflage, as illustrated in figure 8.8.

The goal of camouflage is to foil grouping processes that would normally make the creature stand out from its environment as a separate object. The successfully camouflaged organism is grouped with its surroundings instead, primarily because of the operation of similarity in various guises. If the ani-

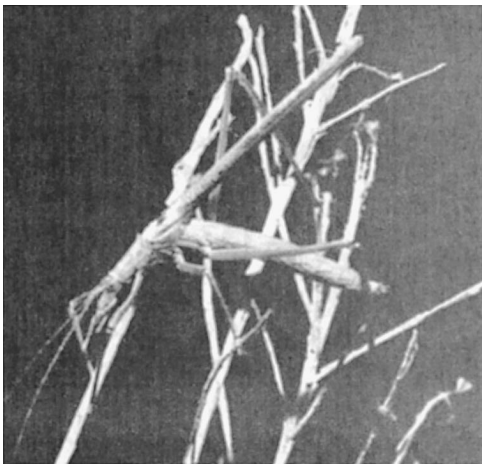


Figure 8.8
An example of natural camouflage. Many animals, birds, and insects exhibit a remarkable ability to blend into their habitual surroundings by foiling many Gestalt principles of grouping. The camouflage is invariably broken when the animal moves relative to the background, however. (Photograph by David C. Rentz.)

mal's coloration and markings are sufficiently similar to its environment in color, orientation, size, and shape, it will be grouped with the background, thus rendering it virtually invisible in the proper context. The effect can be nearly perfect as long as the organism remains stationary, but even perfect camouflage is undone by the principle of common fate once it moves. The common motion of its markings and contours against the background causes them to be strongly grouped together, providing any nearby observer with enough information to perceive it as a separate object.

8.1.3 Measuring Grouping Effects Quantitatively

Gestalt demonstrations of grouping are adequate for establishing the existence of ceteris paribus rules, but they are *not* adequate to support quantitative theories that specify how multiple factors might be integrated. For this purpose, quantitative methods are needed to enable measurement of the amount or degree of grouping. Two such methods have recently been devised, one based directly on reports of grouping and the other based on an indirect but objectively defined task.

Kubovy and Wagemans (1995) measured the relative strength of different groupings by showing observers dot lattices like the one shown in figure 8.9A and measuring the probability with which they reported seeing them organized in various different ways. Such lattices are ambiguous in that they can be seen as being grouped into lines in one of four orientations as indicated in figure 8.9B. Observers were shown a particular lattice for 300 milliseconds (ms) and then were asked to indicate which organization they saw by choosing one among four response symbols representing the possible orientations for that lattice. After many trials, the probabilities of perceiving each grouping could

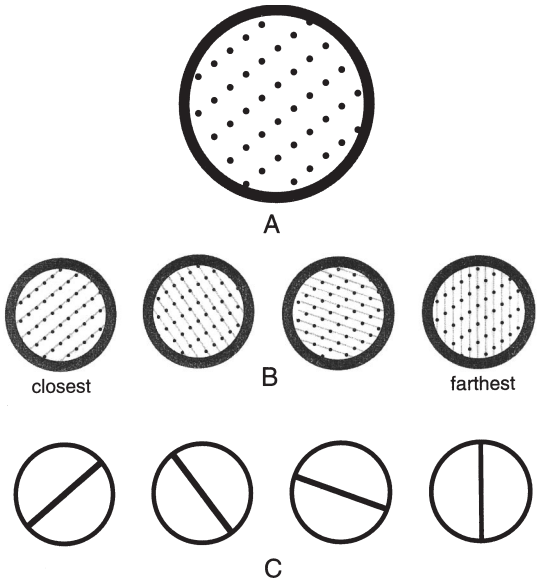


Figure 8.9
Ambiguity in the grouping of dot lattices. Lattices of dots, such as that shown in Part A, can be seen as grouped into lines of different orientations as illustrated in part B by the thin gray lines connecting the dots. Kubovy and Wagemans (1995) had subjects indicate the orientation of dot-lines that they saw by choosing the corresponding response symbol shown in part C. (After Kubovy & Wagemans, 1995.)

be calculated. Consistent with the Gestalt principle of proximity, their results showed that the most likely organization is the one in which the dots are closest together, other organizations being less likely as the spacing between the dots in that orientation increased. Moreover, the data were fit well by a mathematical model in which the attraction between dots decreases exponentially as a function of distance (see also Kubovy, Holcombe, & Wagemans, 1998).

Another quantitative method for studying grouping, called the *repetition discrimination task*, has recently been devised by Palmer and Beck (in preparation). Unlike Kubovy and Wagemans's procedure, this method relies on a task in which there is an objectively correct answer for each response. Subjects are presented with displays like the ones shown in figure 8.10. Each consists of a row of squares and circles that alternate except for a single adjacent pair in which the same shape is repeated. The subject's task on each trial is to determine whether the adjacent repeated pair is composed of squares or circles. They indicate the answer by pressing one button for squares or another for circles as quickly as they can. Response times are measured in three different conditions. In the *within-group* trials, a grouping factor (proximity in figure 8.10A) biases the target pair to be organized into the same group. In the *between-group* trials, the same factor biases the target pair to be organized as part of two different groups (figure 8.10B). In the *neutral* trials, the factor does not bias the pair one way or the other (figure 8.10C). The expectation is that the target pair will be

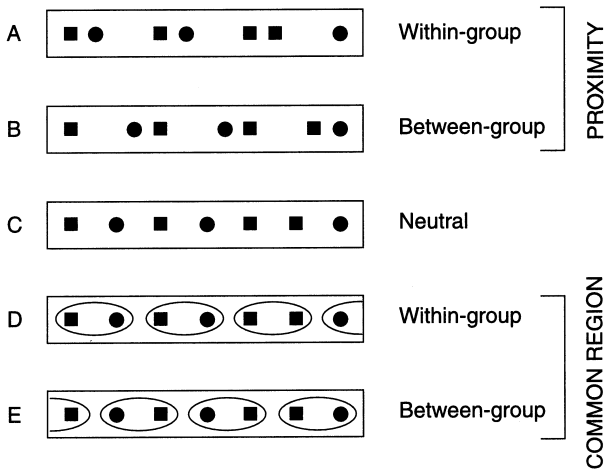


Figure 8.10
Examples of stimuli used in the repetition discrimination task. Subjects must detect whether the adjacent repeated pair are squares or circles. In within-group trials (parts A and D), the repeated elements are within groups defined by a given grouping factor (proximity in part A and common region in part D). In between-group trials (B and E), they are in different groups. In neutral trials (C), no other grouping factor is present.

detected more quickly when it is part of the same group than when it is part of different groups.

The results showed substantial effects of grouping factors on reaction times. Responses were much faster in the within-group trials (719 ms) than in the between-group trials (1144 ms) for the proximity stimuli (figure 8.10A versus figure 8.10B). Responses in the within-group trials were about as fast as those in the neutral trials (730 ms), presumably because the shape similarity of the target pair caused them to be grouped together even in the absence of other grouping factors. Similar results were obtained for detecting adjacent pairs of squares or circles when they were grouped by color similarity, common region, and element connectedness. Figures 8.10C, 8.10D, and 8.10E show neutral, within-group, and between-group displays for the common region experiment. Similar results were obtained, despite the fact that there are no differences in distance between the elements in the target pair. Such findings confirm the importance of grouping factors on this objective perceptual task.

An important advantage of quantitative methods such as these is that they allow precise measurement of grouping effects when phenomenology is unclear. For example, Palmer and Beck used the repetition detection task to determine whether small or large ovals have the greater effect in grouping by common region when they conflict within the same display. Palmer (1992) had previously suggested that smaller regions dominate perception on the basis of the demonstration displays shown in figures 8.11A and 8.11B but admitted that this claim pushed the limits of introspective observations. Using the repetition discrimination task and several stimulus manipulations, Palmer and Beck were able to show that small ovals have a much greater effect than large ovals on response

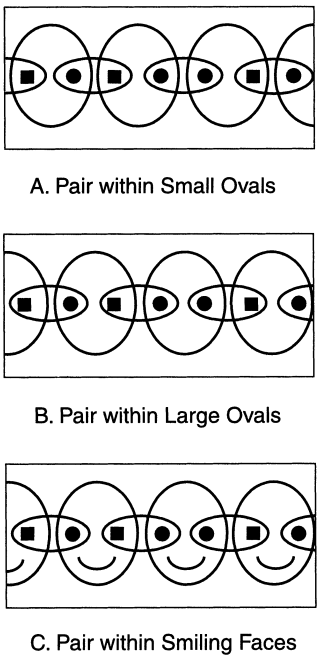


Figure 8.11
Effects of size in common regions. Results from the repetition discrimination task showed that repeated pairs within small regions (A) are detected more quickly than are the same pairs within larger regions (B). This is true even when the large regions were made salient and meaningful by adding "smiles" to form "happy faces."

times in this task and that this difference is due primarily to the size of the ovals rather than their orientation. Somewhat surprisingly, the dominance of the small ovals persisted even when "smiles" were added to the large ovals to make them into faces, as illustrated in figure 8.11C. This finding suggests that grouping in this particular task is not influenced by the familiarity and meaningfulness of faces, which presumably affect perception fairly late in visual processing.

8.1.4 Is Grouping an Early or Late Process?

The question of where in visual processing grouping occurs is an important one. Is it an early process that works at the level of image structure, or does it work later, after depth information has been extracted and perceptual constancy has been achieved? (Recall that perceptual constancy refers to the ability to perceive the unchanging properties of distal environmental objects despite variation in the proximal retinal images caused by differences in viewing conditions.)

Wertheimer (1923/1950) discussed grouping as though it occurred at a very low level, presumably corresponding to what we have called image-based processing. He presented no empirical evidence for this position, but the generally accepted view since his seminal paper has been that organization must

occur early to provide higher-level processes with the perceptual units they require as input. Indeed, this early view has seldom been seriously questioned, at least until recently.

As sensible as the early view of grouping appears a priori, however, there is little empirical evidence to support it. The usual Gestalt demonstrations of grouping do not address this issue because they employ displays in which depth and constancy are irrelevant: two-dimensional displays viewed in the frontal plane with homogeneous illumination. Under these simple conditions it cannot be determined whether the critical grouping factors operate at the level of 2-D image structure or that of 3-D perceptual structure. The reason is that in the Gestalt demonstrations grouping at these two levels—2-D retinal images versus 3-D percepts—lead to the same predictions.

The first well-controlled experiment to explicitly separate the predictions of organization at these two levels concerned grouping by proximity (Rock & Brosgole, 1964). The question was whether the distances that govern proximity grouping are defined in the 2-D image plane or in perceived 3-D space. Rock and Brosgole used a 2-D rectangular array of luminous beads that could be presented to the observer in a dark room either in the frontal plane (perpendicular to the line of sight) or slanted in depth so that the horizontal dimension was foreshortened to a degree that depended on the angle of slant, as illustrated in figure 8.12A. The beads in figure 8.12A were actually closer together vertically, so when they were viewed in the frontal plane, as illustrated in figure 8.12B, observers always reported them as grouped vertically into columns rather than horizontally into rows.

The crucial question was what would happen when the same lattice of beads was presented to the observer slanted in depth so that the beads were closer together horizontally when measured in the retinal image, as depicted in figure 8.12C. (Notice that they are still closer together vertically when measured in the 3-D world.) Not surprisingly, when observers viewed this slanted display with just one eye, so that no binocular depth information was available, they reported the beads to be organized into rows as predicted by retinal proximity. This presumably occurs because they mistakenly perceived the lattice as lying in the frontal plane, even when it was slanted more than 40° in depth. When observers achieved veridical depth perception by viewing the same display binocularly, however, they reported seeing the slanted array of beads as organized into vertical *columns*, just as they did in the frontal viewing condition. This result supports the hypothesis that grouping occurs after stereoscopic depth perception.

Rock, Nijhawan, Palmer, and Tudor (1992) addressed a similar issue in lightness perception: Is similarity grouping by achromatic color based on the retinally measured *luminance* of elements or their phenomenally perceived *lightness* after lightness constancy has been achieved? The first experiment used cast shadows to decouple luminance and lightness. Observers were shown displays similar to the one illustrated in figure 8.13 and were asked to indicate whether the central column of elements grouped with the ones on the left or on the right.

The structure of the display in the critical constancy condition is illustrated in figure 8.13. It was carefully constructed so that the central squares were identi-

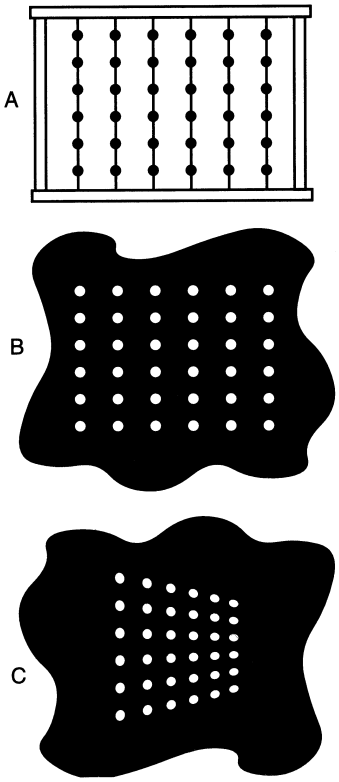


Figure 8.12
Retinal versus perceived distance in proximity grouping. Luminous beads spaced as shown in part A appear to be organized in columns when viewed in the frontal plane (B), because of proximity. When slanted in depth (C) and viewed with both eyes, they are still seen as organized into columns, even when they are closer together horizontally on the retina. This result shows that proximity grouping is influenced by stereoscopic depth processing.

cal in reflectance to the ones on the left (that is, they were made of the same shade of gray paper) but were seen under a shadow cast by an opaque vertical strip hanging nearby. As a result, their luminance—the amount of light reaching the observer's eye after being reflected by the central squares—was identical to the luminance of the squares on the right. Therefore, if grouping were based on relatively early processing of image structure, the central squares would be grouped with the luminance-matched ones on the right. If it were based on relatively late processing after perception of shadows had been achieved, they would group with the reflectance-matched ones on the left. The results showed that grouping followed the predictions of the postconstancy grouping hypothesis: Similarity grouping was governed by the perceived lightness of the squares rather than by their retinal luminance. Other conditions ruled out the possibility that this result was due to simple luminance ratios of the squares to their backgrounds.

Perceptual grouping is also affected by visual completion (Palmer, Neff, & Beck, 1996). Visual completion refers to the fact that when observers see an

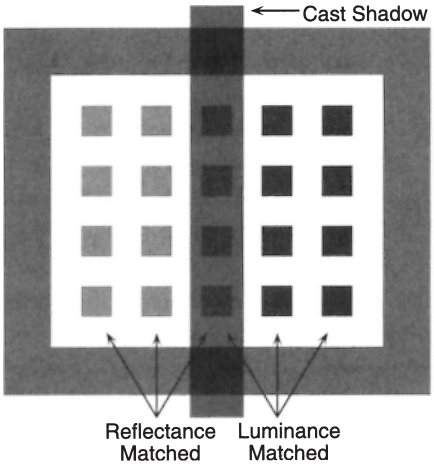


Figure 8.13
Grouping and lightness constancy. When the central column of squares was seen as being in shadow, they were grouped with those of the same reflectance (on the left) rather than those of the same retinal luminance (on the right). This result shows that grouping by achromatic color similarity is influenced by lightness constancy. (After Rock, Nijhawan, Palmer, & Tudor, 1992.)

object partly occluded by another, there is a strong tendency to perceive its shape as being completed behind the occluder. Many theorists believe that this process is relatively late, occurring after perceptual objects and depth relations have already been defined. If grouping by shape similarity is determined by completed shape, this would then be further evidence that it is a relatively late process.

Palmer, Neff, and Beck (1996) investigated whether grouping by shape similarity was determined by the retinal shape of the incomplete elements or by the perceived shape of completed elements. Using the same type of displays as Rock et al. (1992), they constructed a display in which half-circles in the center column were generally perceived as whole circles partly occluded by a vertical strip, as shown in figure 8.14A. An early view of grouping predicts that the central elements will be seen to group with the half-circles on the left because they have the retinal shape of a half-circle. A late view of grouping predicts that they will group with the full circles on the right because they are perceived as being completed behind the occluding strip.

As the reader can see, the central figures group to the right with the completed circles, indicating that grouping is based on similarity of completed shape rather than on retinal shape. The possibility that this outcome was determined by the presence of the occluding strip that divides the elements into two regions according to common region was ruled out by the control condition illustrated in figure 8.14B. Here the occluding strip is simply moved a little further to the side to reveal the entire contour of the central elements, allowing their half-circular shape to be perceived unambiguously. Although common region had a measurable effect in their experiment, most subjects now perceived the central elements as being grouped with the half-circles on the left. These

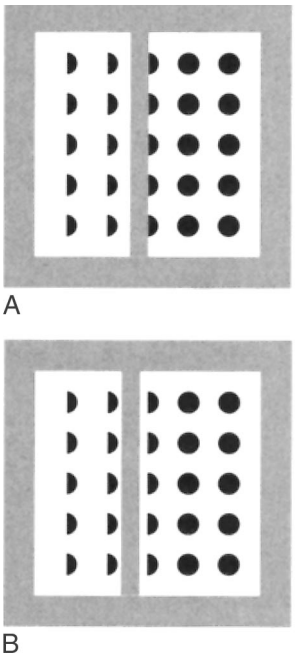


Figure 8.14
Grouping and visual completion. The central column of half circles was grouped more often with the complete ones to their right than with the half-circles to the left when they were seen as partly occluded (A) than when they were seen in their entirety (B). (After Palmer, Neff, & Beck, 1996.)

findings provide further evidence that grouping is a relatively complex and late process in vision.

Such results show that grouping cannot be attributed entirely to early, preconstancy visual processing. However, they are also compatible with the possibility that grouping is a temporally extended process that includes components at both early and later levels of processing. A provisional grouping might be determined at an early, preconstancy stage of image processing but might be overridden if later, object-based information (from depth, lighting conditions, occlusion, and the like) requires it. Evidence that this might be the case could come from cases in which early grouping can be shown to affect constancy operations, in which case grouping must precede constancy processing. Evidence of this sort has not been reported as such in the literature, but this may be because it has not yet been examined rather than because it does not exist. Another sort of evidence for both early and late grouping comes from experiments in which early and late grouping factors combine to produce intermediate results (e.g., Beck, 1975; Olson & Attneave, 1970).

8.1.5 Past Experience

Before we leave the topic of grouping, it is worth pointing out that Wertheimer (1923/1950) discussed one further factor in perceptual grouping that is seldom mentioned: *past experience*. The idea is that if elements have been previously associated in prior viewings, they will tend to be seen as grouped in present

Figure 8.15
Effects of past experience on grouping. Once you se the Dalmatian in the center, it will forever change the grouping you perceive when viewing this picture. This change can only be attributed to past experience, which can have a dramatic effect on perceived organization of ambiguous images. (Photography by R. C. James.)

situations. Figure 8.15 illustrates the point. Initially, you will probably see this picture as a nearly random array of black regions on a white background. Once you are able to see it as a Dalmatian with its head down, sniffing along a street, the picture becomes dramatically reorganized with certain of the dots going together because they are part of the dog and others going together because they are part of the street. The interesting fact is that once you have seen the Dalmatian in this picture, you will continue to see it that way for the rest of your life. Past experience can thus have a dramatic effect on grouping and organization, especially if the organization of the image is highly ambiguous.

The principle of past experience is fundamentally different from the other factors Wertheimer discussed in that it concerns not geometrical properties of the stimulus configuration itself, but rather the viewer's history with respect to the configuration. Perhaps partly for this reason, it has largely been ignored in subsequent presentations of Gestalt principles of grouping. Another reason may be that it is rather easy to show that other grouping factors can block recognition of even the most frequently seen objects (e.g., Gottschaldt, 1929). Figure 8.16 shows an example in which the very simple, common shape of a rectangular prism (figure 8.16A) is completely hidden in a configuration (figure 8.16B) in which good continuation, symmetry, and other intrinsic factors make the embedded prism nearly impossible to perceive. In fairness to past experience, it is important to realize that, unlike the Dalmatian example, in which the

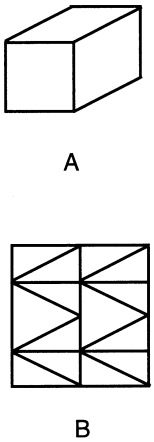


Figure 8.16
Intrinsic grouping factors can overcome past experience. Perception of the familiar shape of a rectangular prism (A) can be blocked by other grouping factors when it is embedded in the context shown in part B.

dots initially appear unorganized, the deck is stacked strongly against seeing the familiar embedded figure by the intrinsic principles of grouping.

One of the reasons the effects of familiarity and object recognition on grouping are theoretically interesting is because they suggest that grouping effects occur as late as object recognition. This should not be too surprising, because the stored representation of the object itself presumably includes information about how its various parts are grouped and related. If part of the object (say, the Dalmatian's head) is identified first, prior knowledge of the shapes of dogs' bodies and legs can be exploited in reorganizing the rest of the image to correspond to these structures. This further process of reorganization suggests that organization is probably occurring *throughout* perception, first at the image-based stage, later at the surface- and object-based stages, and finally at the category-based stage, each result superseding the ones before.

8.2 Region Analysis

The observant reader may have noticed an important gap in the story of perceptual organization as told by the Gestaltists: They neglected to explain how the "elements" of their analysis arise in the first place. Wertheimer appears simply to have assumed the existence of such elements, as though it were so phenomenologically obvious that no analysis was required. If so, this is an example of the very experience error for which the Gestaltists often criticized others. The elements of Wertheimer's displays are not directly given by the structure of the stimulus array, but require an explanation, including an analysis of the factors that govern their existence as perceptual objects.

The obvious basis for the elements of perceptual experience that Wertheimer presupposed in his principles of grouping is an analysis of *regions*: bounded, 2-D areas that constitute spatial subsets of the image. As basic as the concept of a

region is to image processing, we have not yet discussed it explicitly, having concentrated mainly on the essentially one-dimensional constructs of lines and edges. Now we will consider another important aspect of their perceptual function: as boundaries that define 2-D regions. Bounded regions are central to perceptual organization because they may well define the first level of fully 2-D units on which subsequent visual processing is based.

8.2.1 Uniform Connectedness

Palmer and Rock (1994a) provide an explicit analysis of how Wertheimer's presupposed elements might be formed in terms of an organizational principle they call *uniform connectedness*: the tendency to perceive connected regions of uniform image properties—e.g., luminance, color, texture, motion, and disparity—as the initial units of perceptual organization.[1] As we will see, the principle of uniform connectedness also forms a crucial link between the literature on edge detection and that on perceptual organization and grouping.

Let us consider the elements in Wertheimer's original displays as examples of how organization into regions by uniform connectedness might occur as an initial stage in perceptual organization. The dots, lines, and rectangles in figures 8.2A–8.2F are all connected regions of uniform luminance, and they correspond to the elements to which Wertheimer appealed in his analysis of grouping. The V's in figure 8.3A, the lines in figures 8.2G and 8.2H, the X-shaped drawing in figure 8.2I, and the hourglass-shaped contour in figure 8.2J are also uniform connected regions according to Palmer and Rock's analysis, but their relation to Wertheimer's "elements" is slightly more complex and will be considered more fully later.

The powerful effect of uniform connectedness on perceptual organization can be demonstrated in simple displays of dots like those used by Wertheimer, as illustrated in figure 8.17. Part A shows that a row of uniformly spaced dots of different luminance are seen as unitary entities, and part B shows that the same is true for regions that are defined by differently oriented texture elements. Parts C and D show that such regions merge into larger, more complex unitary elements when they are connected by regions defined by the same property, whereas parts E and F show that when they are connected by regions of different properties, they are no longer perceived as fully unitary elements.

One might at first think that uniform connectedness is nothing more than the principle of similarity operating on the basis of luminance and color. For example, if the tiny patch of light falling on each retinal receptor were taken as an element, could uniform connected regions not be explained by grouping these elements according to similarity of luminance and color? Perhaps this is how Wertheimer himself thought about the organization of elements. But sameness of color is not sufficient to explain the perceptual unity of uniform connected regions because it does not account for the difference between *connected* regions of homogeneous color and *disconnected* ones. That is, without the additional constraint of connectedness, there is no basis for predicting that two black areas within the same dot or bar are any more closely related than comparable black areas within two different dots or bars.[2] Phenomenologically speaking, there is no doubt that each individual dot is more tightly organized as a perceptual

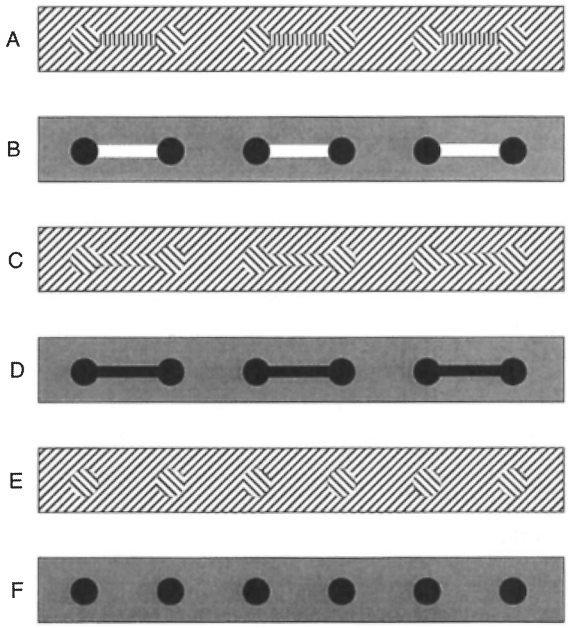


Figure 8.17
Uniform connectedness. Observers perceive connected regions of uniform visual properties as unitary elements whether they are defined by luminance (A and C), texture (parts B and D), or other simple visual properties. Similar elements defined by different properties (E and F) do not have the same unitary nature as those defined by uniform connectedness.

object than is any pair of separate dots. This observation suggests the hypothesis that uniform connectedness is an important principle of perceptual organization.

Palmer and Rock (1994a, 1994b) argue that uniform connectedness cannot be reduced to any principle of grouping because uniform connectedness is not a principle of grouping at all.[3] Their reasoning is that grouping principles presuppose the existence of independent elements that are to be grouped together, whereas uniform connectedness is defined on an unsegregated image. For this reason, uniform connectedness must logically operate *before* any principles of grouping can take effect. This is just another way of saying that because uniform connectedness is the process responsible for forming elements in the first place, it must occur before any process that operates on such elements.

If uniform connectedness is so fundamental in perceptual organization, it is important to understand why. Palmer and Rock argue that it is because of its informational value for designating connected objects (or parts of objects) in the world. As a general rule, if an area of the retinal image constitutes a homogeneous connected region, it almost certainly comes from the light reflected from a single connected object in the environment. This is not invariably true, of course, for the pattern on a camouflaged animal sometimes merges with identically colored regions of the background in its natural habitat, as illustrated in figure 8.8. This is yet another example of a case in which perception goes astray whenever the heuristic assumptions underlying a perceptual process fail to

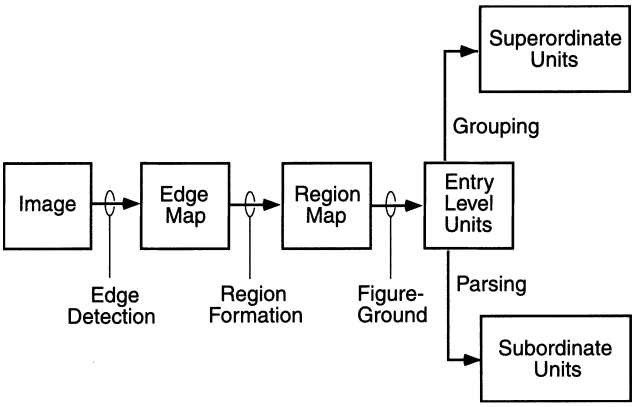


Figure 8.18
A flowchart of Palmer and Rock's (1994a) theory of perceptual organization. After edges are detected, regions are formed, and figure/ground principles operate to form entry-level units. Grouping and parsing can then occur in any order to form higher and lower units in the part/whole hierarchy. (After Palmer & Rock, 1994a.)

hold. Even so, such situations are quite rare, and uniform connectedness is indeed an excellent heuristic for finding image regions corresponding to parts of connected objects in the environment. It therefore makes good sense for the visual system to make a first pass at organizing an image into objects by segregating it into uniform connected regions.

On the basis of this reasoning, Palmer and Rock suggest that uniform connectedness is the first principle of 2-D perceptual organization to operate and the foundation on which all later organization rests. The goal of this initial analysis is to divide the image into a set of mutually exclusive regions—called a *partition* of the image—much like a stained-glass window or a paint-by-numbers template. The regions thus identified can then be further organized by other processes such as discriminating figure from ground, grouping two or more regions together, and parsing a single region into two or more subregions. A flowchart capturing Palmer and Rock's (1994a) view of the relations among these organizational processes is shown in figure 8.18.

Notes

1. Koffka (1935) foreshadowed the idea of uniform connectedness in his discussion of perceptual organization, but he did not examine the implications of his observations, and his brief remarks appear not to have influenced subsequent theories until the rediscovery of the concept by Palmer and Rock (1994a).
2. One other way of accounting for this fact is to appeal to associative grouping (Geisler & Super, in press). The idea of associative grouping is that if A is grouped with B and B is grouped with C, then A will be grouped with C. This hypothesis can be used to explain why the points within a uniform connected region are grouped more strongly with each other than with those of points in other regions.
3. Note that *element* connectedness is a principle of grouping, but *uniform* connectedness is not. These are two different factors of perceptual organization in Palmer and Rock's (1994a) theory that have quite different interpretations.

References

Beck, J. (1975). The relation between similarity grouping and perceptual constancy. *American Journal of Pysychology, 88*(3), 397–409.

Bregman, A. S. (1978). The formation of auditory streams. In J. Requin (Ed.), *Attention and performance* (Vol. 7). Hillsdale, NJ: Erlbaum.

Geisler, W. S., & Super, B. J. (in press). *Psychological Review*.

Goldmeier, E. (1936/1972), Similarity in visually perceived forms. *Psychological Issues, 8*(1), 1–135.

Gottschaldt, K. (1929). Uber den Einfluss der Erfahrung auf die Wahrnemung von Figuren, II. *Psychologische Forschung, 12*, 1–87.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace.

Köhler, W. (1947) *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright.

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology, 35*(1), 71–98.

Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science, 6*(4), 225–234.

Olson, R. K., & Attneave, F. (1970). What variables produce similarity grouping? *American Journal of Psychology, 83*(1), 1–21.

Palmer, S. E. (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology, 24*(3), 436–447.

Palmer, S. E., & Beck, D. (in preparation). The repetition detection task: A quantitative method for studying grouping.

Palmer, S. E., & Levitin, D. (submitted). Synchrony: A new principle of perceptual organization.

Palmer, S. E., & Rock, I. (1994a). On the nature and order of organizational processing: A reply to Peterson. *Psychonomic Bulletin & Review, 1*, 515–519.

Palmer, S. E., & Rock, I. (1994b). Rethinking perceptual organizational: The role of uniform connectedness. *Psychonomic Bulletin & Review, 1*(1), 29–55.

Palmer, S. E., Neff, J., & Beck, D. (1996). Late influences on perceptual grouping: Amodal completion. *Psychonomic Bulletin & Review, 3*(1), 75–80.

Pomerantz, J. R., & Kubovy, M. (1986). Theoretical approaches to perceptual organization: Simplicity and likelihood principles. In K. R. Boff, L. Kaufman, & J. P. Thomas, (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 1–46). New York: Wiley.

Rock, I., Nijhawan, R., Palmer, S., & Tudor, L. (1992). Grouping based on phenomenal similarity of achromatic color. *Perception, 21*(6), 779–789.

Rock, I., & Brosgole, L. (1964). Grouping based on phenomenal proximity. *Journal of Experimental Psychology, 67*, 531–538.

Wertheimer, M. (1923/1950). Untersuchungen zur Lehre von der Gestalt. *Psychologishe Forschung, 4*, 301–350.

Chapter 9

The Auditory Scene

Albert S. Bregman

Historical Difference between Auditory and Visual Perception

If you were to pick up a general textbook on perception written before 1965 and leaf through it, you would not find any great concern with the perceptual or ecological questions about audition. By a perceptual question I mean one that asks how our auditory systems could build a picture of the world around us through their sensitivity to sound, whereas by an ecological one I am referring to one that asks how our environment tends to create and shape the sound around us. (The two kinds of questions are related. Only by being aware of how the sound is created and shaped in the world can we know how to use it to derive the properties of the sound-producing events around us.)

Instead, you would find discussions of such basic auditory qualities as loudness and pitch. For each of these, the textbook might discuss the psychophysical question: which physical property of the sound gives rise to the perceptual quality that we experience? It might also consider the question of how the physiology of the ear and nervous system could respond to those properties of sound. The most perceptual of the topics that you might encounter would be concerned with how the sense of hearing can tell the listener where sounds are coming from. Under this heading, some consideration would be given to the role of audition in telling us about the world around us. For the most part, instead of arising from everyday life, the motivation of much of the research on audition seems to have its origins in the medical study of deafness, where the major concerns are the sensitivity of the auditory system to weak sounds, the growth in perceived intensity with increases in the energy of the signal, and the effects of exposure to noise.

The situation would be quite different in the treatment of vision. It is true that you would see a treatment of psychophysics and physiology, and indeed there would be some consideration of such deficits as colorblindness, but this would not be the whole story. You would also find discussions of higher-level principles of organization, such as those responsible for the constancies. There would, for example, be a description of size constancy, the fact that we tend to see the size of an object as unchanged when it is at a different distance, despite the fact that the image that it projects on our retinas shrinks as it moves further away. Apparently some complex analysis by the brain takes into account clues other than retinal size in arriving at the perceived size of an object.

From chapter 1 in *Auditory Scene Analysis* (Cambridge, MA: MIT Press, 1990), 1–45. Reprinted with permission.

Why should there be such a difference? A proponent of the "great man" theory of history might argue that it was because the fathers of Gestalt psychology, who opened up the whole question of perceptual organization, had focused on vision and never quite got around to audition.

However, it is more likely that there is a deeper reason. We came to know about the puzzles of visual perception through the arts of drawing and painting. The desire for accurate portrayal led to an understanding of the cues for distance and certain facts about projective geometry. This was accompanied by the development of the physical analysis of projected images, and eventually the invention of the camera. Early on, the psychologist was faced with the discrepancy between what was on the photograph or canvas and what the person saw.

The earlier development of sophisticated thinking in the field of visual perception may also have been due to the fact that it was much easier to create a visual display with exactly specified properties than it was to shape sound in equally exact ways. If so, the present-day development of the computer analysis and synthesis of sound ought to greatly accelerate the study of auditory perception.

Of course there is another possibility that explains the slighting of audition in the textbook: Perhaps audition is really a much simpler sense and there are no important perceptual phenomena like the visual constancies to be discovered.

This is a notion that can be rejected. We can show that such complex phenomena as constancies exist in hearing, too. One example is timbre constancy. A friend's voice has the same perceived timbre in a quiet room as at a cocktail party. Yet at the party, the set of frequency components arising from that voice is mixed at the listener's ear with frequency components from other sources. The total spectrum of energy that reaches the ear may be quite different in different environments. To recognize the unique timbre of the voice we have to isolate the frequency components that are responsible for it from others that are present at the same time. A wrong choice of frequency components would change the perceived timbre of the voice. The fact that we can usually recognize the timbre implies that we regularly choose the right components in different contexts. Just as in the case of the visual constancies, timbre constancy will have to be explained in terms of a complicated analysis by the brain, and not merely in terms of a simple registration of the input by the brain.

There are some practical reasons for trying to understand this constancy. There are engineers currently trying to design computers that can understand what a person is saying. However, in a noisy environment the speaker's voice comes mixed with other sounds. To a naive computer, each different sound that the voice comes mixed with makes it sound as if different words were being spoken or as if they were spoken by a different person. The machine cannot correct for the particular listening conditions as a human can. If the study of human audition were able to lay bare the principles that govern the human skill, there is some hope that a computer could be designed to mimic it.

The Problem of Scene Analysis

It is not entirely true that textbooks ignore complex perceptual phenomena in audition. However, they are often presented as an array of baffling illusions.[1]

AI CSAITT STIOTOS

A_I $C_S A_I T_T$ $S_T I_O T_O S$

Figure 9.1
Top line: a string of letters that makes no sense because it is a mixture of two messages. Bottom line: the component messages are segregated by visual factors. (From Bregman 1981.)

They seem more like disconnected fragments than a foundation for a theory of auditory perception. My purpose in this book is to try to see them as oblique glimpses of a general auditory process of organization that has evolved, in our auditory systems, to solve a problem that I will refer to as "auditory scene analysis."

Let me clarify what I mean by auditory scene analysis. The best way to begin is to ask ourselves what perception is for. Since Aristotle, many philosophers and psychologists have believed that perception is the process of using the information provided by our senses to form mental representations of the world around us. In using the word representations, we are implying the existence of a two-part system: one part forms the representations and another uses them to do such things as calculate appropriate plans and actions. The job of perception, then, is to take the sensory input and to derive a useful representation of reality from it.

An important part of building a representation is to decide which parts of the sensory stimulation are telling us about the same environmental object or event. Unless we put the right combination of sensory evidence together, we will not be able to recognize what is going on. A simple example is shown in the top line of figure 9.1. The pattern of letters is meaningful, but the meaning cannot be extracted because the letters are actually a mixture from two sentences, and the two cannot be separated. However, if, as in the lower line of the figure, we give the eyes some assistance, the meaning becomes apparent.

This business of separating evidence has been faced in the design of computer systems for recognizing the objects in natural scenes or in drawings. Figure 9.2 shows a line drawing of some blocks.[2] We can imagine that the picture has been translated into a pattern in the memory of the computer by some process that need not concern us. We might think that once it was entered, all that we would have to do to enable the computer to decide which objects were present in the scene would be to supply it with a description of the shape of each possible one. But the problem is not as easy as all that. Before the machine could make any decision, it would have to be able to tell which parts of the picture represented parts of the same object. To our human eyes it appears that the regions labeled A and B are parts of a single block. This is not immediately obvious to a computer. In simple line drawings there is a rule that states that any white area totally surrounded by lines must depict a single surface. This rule implies that in figure 9.2 the whole of region A is part of a single surface. The reason for grouping region A with B is much more complex. The question of how it can be done can be set aside for the moment. The point of the example is that unless regions A and B are indeed considered part of a single object, the description that the computer will be able to construct will not be correct and

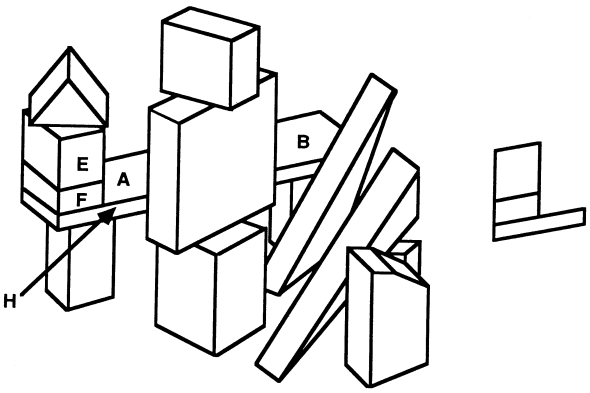


Figure 9.2
A line drawing of blocks for visual scene analysis. (After Guzman 1969.)

the elongated shape formed out of A, B, and other regions will not be seen. It seems as though a preliminary step along the road to recognition would be to program the computer to do the equivalent of taking a set of crayons and coloring in, with the same color, all those regions that were parts of the same block. Then some subsequent recognition process could simply try to form a description of a single shape from each set in which the regions were the same color. This allocation of regions to objects is what is known to researchers in machine vision as the scene analysis problem.

There are similar problems in hearing. Take the case of a baby being spoken to by her mother. The baby starts to imitate her mother's voice. However, she does not insert into the imitation the squeaks of her cradle that have been occurring at the same time. Why not? A physical record of what she has heard would include them. Somehow she has been able to reject the squeak as not being part of the perceptual "object" formed by her mother's voice. In doing so, the infant has solved a scene analysis problem in audition.

It is important to emphasize again that the way that sensory inputs are grouped by our nervous systems determines the patterns that we perceive. In the case of the drawings of blocks, if areas E, F, and H were grouped as parts of the same object, we would see the L-shaped object shown at the right. The shape of the object formed by this grouping of areas is an emergent property, since it is not a property of any of the parts taken individually, but emerges only as a result of the grouping of the areas. Normally, in perception, emergent properties are accurate portrayals of the properties of the objects in our environment. However, if scene analysis processes fail, the emergent perceived shapes will not correspond to any environmental shapes. They will be entirely chimerical.

The difficulties that are involved in the scene analysis processes in audition often escape our notice. This example can make them more obvious. Imagine that you are on the edge of a lake and a friend challenges you to play a game. The game is this: Your friend digs two narrow channels up from the side of the lake. Each is a few feet long and a few inches wide and they are spaced a few feet apart. Halfway up each one, your friend stretches a handkerchief and fas-

tens it to the sides of the channel. As waves reach the side of the lake they travel up the channels and cause the two handkerchiefs to go into motion. You are allowed to look only at the handkerchiefs and from their motions to answer a series of questions: How many boats are there on the lake and where are they? Which is the most powerful one? Which one is closer? Is the wind blowing? Has any large object been dropped suddenly into the lake?

Solving this problem seems impossible, but it is a strict analogy to the problem faced by our auditory systems. The lake represents the lake of air that surrounds us. The two channels are our two ear canals, and the handkerchiefs are our ear drums. The only information that the auditory system has available to it, or ever will have, is the vibrations of these two ear drums. Yet it seems to be able to answer questions very like the ones that were asked by the side of the lake: How many people are talking? Which one is louder, or closer? Is there a machine humming in the background? We are not surprised when our sense of hearing succeeds in answering these questions any more than we are when our eye, looking at the handkerchiefs, fails.

The difficulty in the examples of the lake, the infant, the sequence of letters, and the block drawings is that the evidence arising from each distinct physical cause in the environment is compounded with the effects of the other ones when it reaches the sense organ. If correct perceptual representations of the world are to be formed, the evidence must be partitioned appropriately.

In vision, you can describe the problem of scene analysis in terms of the correct grouping of regions. Most people know that the retina of the eye acts something like a sensitive photographic film and that it records, in the form of neural impulses, the "image" that has been written onto it by the light. This image has regions. Therefore, it is possible to imagine some process that groups them. But what about the sense of hearing? What are the basic parts that must be grouped to make a sound?

Rather than considering this question in terms of a direct discussion of the auditory system, it will be simpler to introduce the topic by looking at a spectrogram, a widely used description of sound. Figure 9.3 shows one for the spoken word "shoe." The picture is rather like a sheet of music. Time proceeds from left to right, and the vertical dimension represents the physical dimension of frequency, which corresponds to our impression of the highness of the sound. The sound of a voice is complex. At any moment of time, the spectrogram shows more than one frequency. It does so because any complex sound can actually be viewed as a set of simultaneous frequency components. A steady pure tone, which is much simpler than a voice, would simply be shown as a horizontal line because at any moment it would have only one frequency.

Once we see that the sound can be made into a picture, we are tempted to believe that such a picture could be used by a computer to recognize speech sounds. Different classes of speech sounds, stop consonants such as "b" and fricatives such as "s" for example, have characteristically different appearances on the spectrogram. We ought to be able to equip the computer with a set of tests with which to examine such a picture and to determine whether the shape representing a particular speech sound is present in the image. This makes the problem sound much like the one faced by vision in recognizing the blocks in figure 9.2.

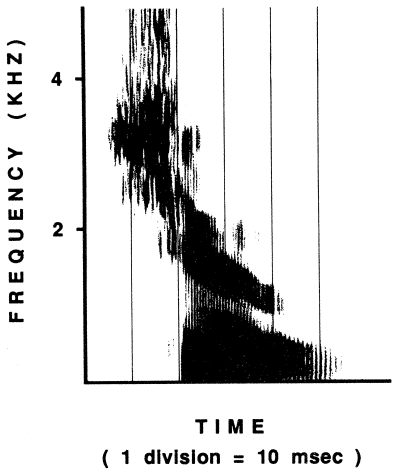


Figure 9.3
Spectrogram of the word "shoe" spoken in isolation.

If a computer could solve the recognition problem by the use of a spectrogram, it would be very exciting news for researchers in human audition, because there is some reason to believe that the human auditory system provides the brain with a pattern of neural excitation that is very much like a spectrogram. Without going into too much detail, we can sketch this process as follows. As sound enters the ear, it eventually reaches a part called the inner ear where it affects an organ called the basilar membrane, a long coiled ribbon. Different frequency components in the incoming sound will cause different parts of this organ to vibrate most vigorously. It reacts most strongly to the lowest audible frequencies at one end, to the highest at the other, with an orderly progression from low to high in between. A different group of neurons connects with each location along the basilar membrane and is responsible for recording the vibration at that location (primarily). As the sound changes over time, different combinations of neural groups are activated. If we imagined the basilar membrane oriented vertically so that the neural groups responsive to the highest frequencies were at the top, and also imagined that each group was attached to a pen, with the pen active whenever a neural group was, the pens would write out a picture of the sound that looked like a spectrogram. So the brain has all the information that is visible in the spectrogram, and providing that it could store a record of this information for some brief period of time, it would have a neural spectrogram.

The account that I have just given hides a deep problem. The spectrographic record of most situations would not have the pristine purity of figure 9.3, which represents speech recorded in an absolutely quiet background. The real world is a great deal messier. A typical acoustic result is shown in figure 9.4. Here all the sounds are being mixed together in the listener's ear in exactly the same way that the waves of the lake, in our earlier example, were mixed in each of the channels that ran off it. The spectrogram for a mixture of sounds looks some-

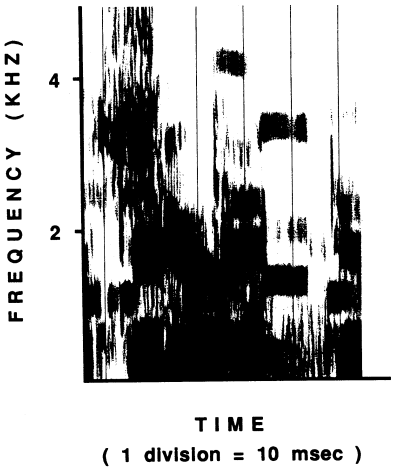


Figure 9.4
A spectrogram of a mixture of sounds (containing the word "shoe").

what like a picture created by making a spectrogram of each of the individual sounds on a separate piece of transparent plastic, and then overlaying the individual spectrograms to create a composite. The spectrogram of the word shoe is actually one of the component spectrograms of the mixture.

Although the theorist has the privilege of building the composite up from the pictures of its components, the auditory system, or any machine trying to imitate it, would be presented only with the spectrogram of the mixture and would have to try to infer the set of pictures that was overlaid to produce it.

The recognizer would have to solve the following problems: How many sources have created the mixture? Is a particular discontinuity in the picture a change in one sound or an interruption by a second one? Should two dark regions, one above the other in the picture (in other words, occurring at the same time), be grouped as a single sound with a complex timbre or separated to represent two simultaneous sounds with simpler timbres? We can see that if we look at a spectrogram representing a slice of real life, we would see a complex pattern of streaks, any pair of which could have been caused by the same acoustic event or by different ones. A single streak could have been the summation of one, two, or even more parts of different sounds. Furthermore, the frequency components from one source could be interlaced with those of another one; just because one horizontal streak happens to be immediately above another, it does not mean that they both arose from the same sonic event.

We can see that just as in the visual problem of recognizing a picture of blocks, there is a serious need for regions to be grouped appropriately. Again, it would be convenient to be able to hand the spectrogram over to a machine that did the equivalent of taking a set of crayons and coloring in, with the same color, all the regions on the spectrogram that came from the same source. This "coloring problem" or "auditory scene analysis problem" is what the rest of this chapter is about.

Objects Compared to Streams

It is also about the concept of "auditory streams." An auditory stream is our perceptual grouping of the parts of the neural spectrogram that go together. To see the reasons for bringing in this concept, it is necessary to consider the relations between the physical world and our mental representations of it. As we saw before, the goal of scene analysis is the recovery of separate descriptions of each separate thing in the environment. What are these things? In vision, we are focused on objects. Light is reflected off objects, bounces back and forth between them, and eventually some of it reaches our eyes. Our visual sense uses this light to form separate descriptions of the individual objects. These descriptions include the object's shape, size, distance, coloring, and so on.

Then what sort of information is conveyed by sound? Sound is created when things of various types happen. The wind blows, an animal scurries through a clearing, the fire burns, a person calls. Acoustic information, therefore, tells us about physical "happenings." Many happenings go on at the same time in the world, each one a distinct event. If we are to react to them as distinct, there has to be a level of mental description in which there are separate representations of the individual ones.

I refer to the perceptual unit that represents a single happening as an auditory stream. Why not just call it a sound? There are two reasons why the word stream is better. First of all a physical happening (and correspondingly its mental representation) can incorporate more than one sound, just as a visual object can have more than one region. A series of footsteps, for instance, can form a single experienced event, despite the fact that each footstep is a separate sound. A soprano singing with a piano accompaniment is also heard as a coherent happening, despite being composed of distinct sounds (notes). Furthermore, the singer and piano together form a perceptual entity—the "performance"—that is distinct from other sounds that are occurring. Therefore, our mental representations of acoustic events can be multifold in a way that the mere word "sound" does not suggest. By coining a new word, "stream," we are free to load it up with whatever theoretical properties seem appropriate.

A second reason for preferring the word "stream" is that the word "sound" refers indifferently to the physical sound in the world and to our mental experience of it. It is useful to reserve the word "stream" for a perceptual representation, and the phrase "acoustic event" or the word "sound" for the physical cause.

I view a stream as a computational stage on the way to the full description of an auditory event. The stream serves the purpose of clustering related qualities. By doing so, it acts as a center for our description of an acoustic event. By way of analogy, consider how we talk about visible things. In our verbal descriptions of what we see, we say that an *object* is red, or that it is moving fast, that it is near, or that it is dangerous. In other words, the notion of an object, understood whenever the word "it" occurs in the previous sentence, serves as a center around which our verbal descriptions are clustered. This is not just a convenience of language. The perceptual representation of an object serves the same purpose as the "it" in the sentence. We can observe this when we dream. When, for some reason, the ideas of angry and dog and green are pulled out

from our memories, they tend to coalesce into a single entity and we experience an angry green dog and not merely anger, greenness, and dogness taken separately. Although the combination of these qualities has never occurred in our experience, and therefore the individual qualities must have been dredged up from separate experiences, those qualities can be experienced visually only as properties of an *object*. It is this "belonging to an object" that holds them together.

The stream plays the same role in auditory mental experience as the object does in visual. When we want to talk about auditory units (the auditory counterparts of visual objects), we generally employ the word "sound." We say that a sound is high pitched or low, that it is rising or falling, that it is rough or smooth, and so on. Again I am convinced that this is not simply a trick of language, but an essential aspect of both our conceptual and our perceptual representations of the world. Properties have to belong to something. This becomes particularly important when there is more than one "something" in our experience. Suppose there are two acoustic sources of sound, one high and near and the other low and far. It is only because of the fact that nearness and highness are grouped as properties of one stream and farness and lowness as properties of the other that we can experience the uniqueness of the two individual sounds rather than a mush of four properties.

A critic of this argument might reply that the world itself groups the "high" with the "near" and the "low" with the "far." It is not necessary for us to do it. However, it is not sufficient that these clusters of properties be distinct in the physical happenings around us. They must also be assigned by our brains to distinct mental entities. In auditory experience, these entities are the things that I am calling streams. As with our visual experience of objects, our auditory streams are ways of putting the sensory information together. This going together has obvious implications for action. For example, if we assign the properties "far" and "lion roar" to one auditory stream and the properties "near" and "crackling fire" to another one, we might be inclined to behave differently than if the distance assignments had been reversed.

When people familiar with the English language read the phrase "The gray wagon was on the black road," they know immediately that it is the wagon that is gray, not the road. They know it because they can *parse* the sentence, using their knowledge of the English syntax to determine the correct "belongingness" relations between the concepts. Similarly, when listeners create a mental representation of the auditory input, they too must employ rules about what goes with what. In some sense, they can be said to be parsing this input too.

The Principle of Exclusive Allocation

Any system that attempts to build descriptions of a natural world scene must assign the perceptual qualities that it creates to one organization or another. The quality "loud" is assigned to the organization that represents the roar of the lion. The quality "far" is assigned as the distance of that same event. The Gestalt psychologists made this point by introducing the principle of belongingness. In describing the visual organization of drawings like the one in figure 9.5, they pointed out that the lines at which the drawn irregular figure overlaps the circle (shown as a dark line in part B of the figure) are generally seen as part

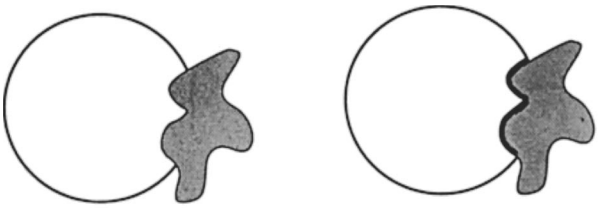


Figure 9.5
An example of "belongingness." The dark portion of the line seems to belong to the irregular form.

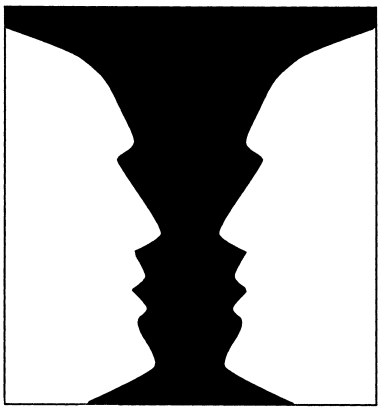


Figure 9.6
An ambiguous drawing in which either a vase at the center or two faces at the sides can be seen.

of the irregular figure and not of the circle. That is, they *belong* to the irregular form. With an effort, we can see them as part of a circle; then they belong to the circle. In any mental representation of a drawing, a perceived line always belongs to some figure of which it forms a part. The belongingness may shift, for example, when we try to see the figure in a different way, but regardless of how we see it, it is always a property *of* something.

There is a second principle that I want to introduce here because it has a connection with the principle of belongingness. This is the principle of "exclusive allocation." It can be seen in an ambiguous visual figure such as the vase-faces illusion of the Gestalt psychologists. An example is shown in figure 9.6. We can interpret the figure as an outline of either a vase or two faces. The "exclusive allocation of evidence" describes how these interpretations affect the line that separates the vase from a face. When we see the vase, that line is allocated to the vase and defines its shape. When we see the face, the same line is now allocated to the face. It is never allocated to both vase and face at the same time, but exclusively to one of them.

The exclusive allocation principle says that a sensory element should not be used in more than one description at a time. If the line is assigned to the vase, that assignment "uses up" the line so that its shape cannot contribute to the shape of another figure at the same time. There are certain limits to this idea,

but it holds true often enough that it is worth pointing it out as a separate principle. It is not identical to the principle of belongingness. The latter merely states that the line has to be seen as a property of a figure, but does not prevent it from being allocated to more than one at a time.

There is a certain ecological validity of the principle of exclusive allocation in vision. The term "ecological validity" means that it tends to give the right answers about how the visual image has probably originated in the external world. In the case of edges separating objects, there is a very low likelihood (except in jigsaw puzzles) that the touching edges of two objects will have the same shape exactly. Therefore the shape of the contour that separates our view of two objects probably tells us about the shape of only one of them—the nearer one. The decision as to which object the contour belongs to is determined by a number of cues that help the viewer to judge which object is closer.

Dividing evidence between distinct perceptual entities (visual objects or auditory streams) is useful because there really are distinct physical objects and events in the world that we humans inhabit. Therefore the evidence that is obtained by our senses really ought to be untangled and assigned to one or another of them.

Our initial example came from vision, but the arguments in audition are similar. For example, it is very unlikely that a sound will terminate at exactly the moment that another begins. Therefore when the spectral composition of the incoming sensory data changes suddenly, the auditory system can conclude that only one sound in a mixture has gone on or off. This conclusion can give rise to a search in the second sound for a continuation of the first one.

The strategy completes itself in the following way. Let us give the name A to the segment of sound that occurs prior to the change, and call the second part B. If spectral components are found in B that match the spectrum of A, they are considered to be the continuing parts of A. Accordingly, they can be subtracted out of B. This allows us a picture of the second sound free from the influence of the first. This is called the "old-plus-new heuristic," and it is shown to be one of our most powerful tools in solving the scene analysis problem in audition. Here I want to point out that it is an example of the principle of exclusive allocation in which the allocation of the continuing spectral components to the first sound interferes with their being allocated to the second.

Another case of exclusive allocation is shown in an experiment by Bregman and Rudnicky, using the pattern of pure tones shown in figure 9.7.[3] In this figure the horizontal dimension represents time and the vertical one shows the frequency of the tones. The listener's task was to decide on the order of two target tones, A and B, embedded in the sequence. Were they in the order high-low or low-high? When A and B were presented alone, as an isolated pair of tones, this decision was very easy. However, when the two tones labeled F (for "flankers") were added to the pattern, the order of A and B became very hard to hear. Apparently when they were absorbed as the middle elements of a larger pattern, FABF, the orders AB and BA lost their uniqueness.

This experiment was about the perceptual allocation of the F tones. As long as they were allocated to the same auditory stream as A and B, the order of A and B was hard to hear. However, Bregman and Rudnicky reasoned that if

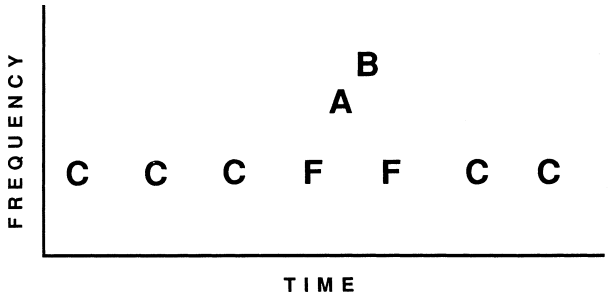


Figure 9.7
A tone sequence of the type used by Bregman and Rudnicky (1975).

some principle of grouping were able to assign the F tones to a different perceptual stream, the order of A and B might become audible again. With this in mind, they introduced yet another group of tones, labeled C (for "captors") in figure 9.7. They varied the frequency of these C tones. When they were very low, much lower than the frequency of the F tones, the F tones grouped with the AB tones and the order of A and B was unclear to the listeners. However, when the C tones were brought up close to the frequency of the F tones, they captured them into a stream, CCCFFCC. One reason for this capturing is that tones tend to group perceptually with those that are nearest to them in frequency; a second is that the F tones were spaced so that they fell into a regular rhythmic pattern with the C tones. When the capturing occurred, the order of AB was heard more clearly because they were now in their own auditory stream that was separate from the CCCFCC stream. The belongingness of the F tones had been altered, and the perceived auditory forms were changed.

Scene analysis, as I have described it, involves putting evidence together into a structure. Demonstrations of the perceptual systems acting in this way are seen in certain kinds of illusions where it appears that the correct features of the sensory input have been detected but have not been put together correctly. Two examples will make this clearer.

The first is in vision. Treisman and Schmidt carried out an experiment in which a row of symbols was flashed briefly in a tachistoscope.[4] There were three colored letters flanked by two black digits. The viewers were asked to first report what the digits were and then to report on the letters. Their reports of the digits were generally correct, but the properties of the letters were often scrambled. A subject might report a red O and a green X, when actually a green O and a red X had been presented. These combinations of features often seemed to the viewers to be their actual experiences rather than merely guesses based on partially registered features of the display. The experimenters argued that this showed that the human mind cannot consciously experience disembodied features and must assign them to perceived objects. That is, the mind obeys the principle of belongingness.

The second example comes from audition. In 1974, Diana Deutsch reported an interesting illusion that could be created when tones were sent to both ears of a listener over headphones. The listener was presented with a continuously

repeating alternation of two events. Event A was a low tone presented to the left ear, accompanied by a high tone presented to the right ear. Event B was just the reverse: a low tone to the right ear together with a high tone to the left. The high and low tones were pure sine wave tones spaced exactly an octave apart. Because events A and B alternated, each ear was presented with a sequence of high and low tones. Another way to express it is that while both the high and low tones bounced back and forth between the ears, the high and low were always in opposite ears.

However the experience of many listeners did not resemble this description. Instead they heard a single sound bouncing back and forth between the ears. Furthermore, the perceived tone alternated between sounding high pitched and sounding low as it bounced from side to side. The only way this illusion could be explained was to argue that the listeners were assuming the existence of a single tone, deriving two different descriptions of it from two different types of perceptual analyses, and then putting the two descriptions together incorrectly. Apparently they derived the fact that the tone was changing in frequency by monitoring the changes in a single ear (usually the right). However, they derived the *position* of the assumed single sound by tracking the position of the higher tone. Therefore, they might report hearing a low tone on the left at the point in time at which, in actuality, a high tone had been presented on the left. Here we see an example of pitch and location assigned in the wrong combination to the representation of a sound. Therefore, this can be classified as a misassignment illusion just as Treisman and Schmidt's visual illusion was.

The question of why this illusion occurs can be set aside for the moment. What is important is that the illusion suggests that an assignment process is taking place, and this supports the idea that perception is a process of building descriptions. Only by being built could they be built incorrectly.

These illusions show that there are some similarities in how visual and auditory experiences are organized. A thoughtful discussion of the similarities and differences between vision and audition can be found in a paper by Bela Julesz and Ira Hirsh.[5] There is no shortage of parallels in audition to visual processes of organization. This chapter cannot afford the space to mention many examples, but it can at least discuss two of them, the streaming phenomenon and the continuity illusion.

Two Comparisons of Scene Analysis in Vision and Audition

Auditory Streaming and Apparent Motion

One auditory phenomenon with a direct parallel in vision is the auditory streaming effect. This is the phenomenon that originally got me interested in auditory organization. The effect occurred when listeners were presented with an endlessly repeating loop of tape on which were recorded a sequence of six different tones, three high ones and three low ones. The high ones were at least one and a half octaves above the low ones. High and low tones alternated. If tones are given numbers according to their pitches with 1 as the lowest and 6 as the highest the tones were arranged in the sequence 142536. The six tones, shown in figure 9.8, formed a repeating loop that was cycled over and over.

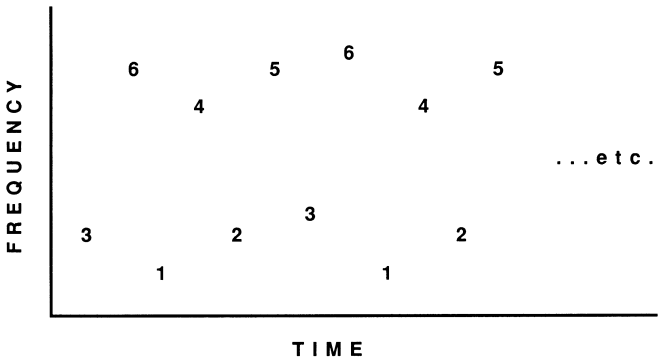


Figure 9.8
A repeating cycle of six tones, of the type used by Bregman and Campbell (1971).

When the cycle of tones was presented very slowly the listeners heard the sequence of high and low tones in the order in which they occurred on the tape. However, as it was made faster, a strange perceptual effect became stronger and stronger and was extremely compelling when there was only one-tenth of a second between the onsets of consecutive tones. When the effect occurred, the listeners did not actually hear the tones in the correct order, 142536. Instead, they heard two streams of tones, one containing a repeating cycle of the three low pitched tones, 1–2–3– (where dashes indicate silences) and the other containing the three high ones (–4–5–6). The single sequence of tones seemed to have broken up perceptually into two parallel sequences, as if two different instruments were playing different, but interwoven parts. Furthermore it was impossible for the listeners to focus their attention on both streams at the same time. When they focused on one of the streams, the other was heard as a vague background. As a consequence, while the listeners could easily judge the order of the high tones taken alone, or of the low ones taken alone, they could not put this information together to report the order of the six tones in the loop. Many listeners actually reported that the high tones all preceded the low ones, or vice versa, although this was never the case.

Other research has shown that the phenomenon of stream segregation obeys some fairly simple laws. If there are two sets of tones, one of them high in frequency and the other low, and the order of the two sets is shuffled together in the sequence (not necessarily a strict alternation of high and low), the degree of perceptual segregation of the high tones from the low ones will depend on the frequency separation of the two sets. Therefore if the two conditions shown in figure 9.9 are compared, the one on the right will show greater perceptual segregation into two streams. An interesting point is that visually, looking at figure 9.9, the perception of two distinct groups is also stronger on the right.

There is another important fact about stream segregation: the faster the sequence is presented, the greater is the perceptual segregation of high and low tones. Again there is a visual analogy, as shown in figure 9.10. We see the pattern in the right panel, in which there is a contraction of time (the same as an increase in speed), as more tightly grouped into two groups than the left panel is.

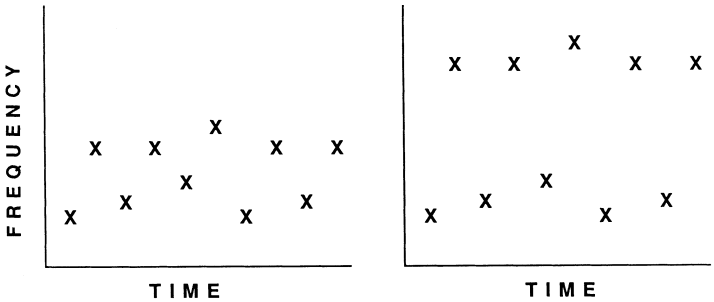


Figure 9.9
Stream segregation is stronger when the frequency separation between high and low tones is greater, as shown on the right.

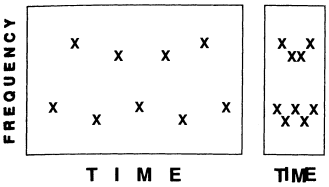


Figure 9.10
Stream segregation is higher at higher speeds, as shown on the right.

Gestalt Grouping Explanation

In the visual analogies, the grouping is predictable from the Gestalt psychologists' proximity principle, which states roughly that the closer the visual elements in a set are to one another, the more strongly we tend to group them perceptually. The Gestalt psychologists thought of this grouping as if the perceptual elements—for example, the notes in figure 9.9—were attracting one another like miniature planets in space with the result that they tended to form clusters in our experience. If the analogy to audition is a valid one, this suggests that the spatial dimension of distance in vision has two analogies in audition. One is separation in time, and the other is separation in frequency. Both, according to this analogy, are distances, and Gestalt principles that involve distance should be valid for them.

The Gestalt principles of grouping were evolved by a group of German psychologists in the early part of this century to explain why elements in visual experience seemed highly connected to one another despite the fact that the incoming light rays, pressure energy, sound waves, and so on stimulated discrete sensory receptors such as the ones found in the retina of the eye. The word Gestalt means "pattern" and the theory described how the brain created mental patterns by forming connections between the elements of sensory input. We cannot go into much detail here about this subtle and philosophically sophisticated theory. However, we can examine a few of the observations that they made about the grouping of sensory elements. They are illustrated in the present discussion by means of the set of diagrams shown in figure 9.11.

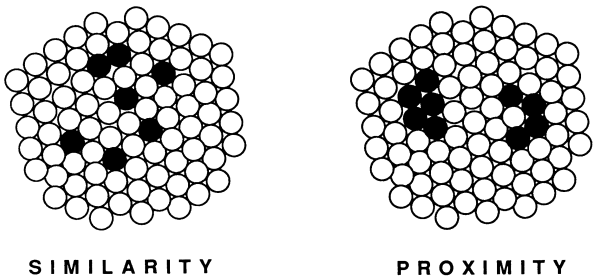


Figure 9.11
Illustration of the effects of the Gestalt principles of similarity and proximity on visual grouping.

Distinct visible elements will be grouped to form coherent perceptual organizations if they fulfill certain conditions. The first is similarity. In the first part of the figure, the black and white blobs can be seen as different subgroups because of the similarity of color within each group and the contrast between groups. Similarly, in audition we find that sounds of similar timbres will group together so that the successive sounds of the oboe will segregate from those of the harp, even when they are playing in the same register.

The second part of the figure shows grouping by a second factor, proximity, where the black blobs seem to fall into two separate clusters because the members of one cluster are closer to other members of the same one than they are to the elements that form the other one. It would appear then that the example of stream segregation would follow directly from the Gestalt law of grouping by proximity. The high tones are closer to one another (in frequency) than they are to the low ones. As the high and low groups are moved further away from one another in frequency, the within-group attractions will become much stronger than the between-group attractions. Speeding the sequence up simply has the effect of moving things closer together on the time dimension. This attenuates the differences in time separations and therefore reduces the contribution of separations along the time dimension to the overall separation of the elements. In doing so, it exaggerates the effects of differences in the frequency dimension, since the latter become the dominant contributors to the total distance.

In both parts of figure 9.11, it is not just that the members of the same group go with one another well. The important thing is that they go with one another *better* than they go with members of the other group. The Gestalt theorists argued that there was always competition between the "forces of attraction" of elements for one another and that the perceptual organization that came out of this conflict would be a consequence of the distribution of forces across the whole perceptual "field," and not of the properties of individual parts taken in isolation.

The Gestalt psychologists' view was that the tendency to form perceptual organizations was innate and occurred automatically whenever we perceived anything. It was impossible, they claimed, to perceive sensory elements without their forming an organized whole. They argued that this organizing tendency was an automatic tendency of brain tissue.

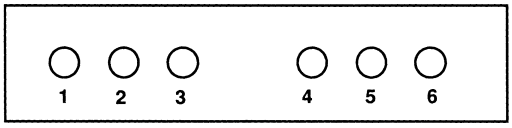


Figure 9.12
A visual display used to demonstrate visual motion segregation. Two groups of three lamps are arranged in a horizontal row.

Auditory Streaming versus Apparent Motion

We have been examining the phenomenon of auditory stream segregation as an example of how phenomena of auditory organization can exhibit the same complexities as are found in vision. This has led us to see interesting parallels in the principles that govern auditory stream segregation and visual grouping. But we have not yet discussed the most striking parallel, that between auditory stream segregation and the phenomenon of apparent motion in vision. Apparent motion is the perceptual effect that used to be very popular on the billboards of theatres, where the switching on and off of a series of electric light bulbs in sequence gave the experience of movement. In the laboratory it is usually created in a much simpler form. Two electric lamps, often seen as small white dots in an otherwise black room, are alternately switched on, each for a brief instant, so that a movement is seen that dances back and forth between the lights, always moving from the light that has just been flashed to the light that is currently being flashed. If the lamps are close together, it may seem that the light itself is moving back and forth. At greater distances the experience is just an impression of movement.

In 1915, Körte formulated a number of laws relating the duration, brightness, and spatial separation of the lamps to the strength of the impression of movement. Körte's third law stated that within certain ranges, if you want to increase the spatial separation between the lamps and still have a strong impression of motion, you had to slow down the alternation of flashes. It was almost as if the movement would not be able to keep up with the alternation of flashes if they were far separated in space unless the flashes were slowed down to compensate for their separation.

A more elaborate form of the apparent motion effect strongly resembles the streaming effect.[6] Instead of two lamps, there are six, arranged in a horizontal row as shown in figure 9.12. They are arranged so that there is a wider gap between the left triplet of lights and the right triplet than there is between the lights within each triplet. If we label the lamps with the digits 1 to 6 from left to right, the order in which the lights are to be flashed can be expressed as the sequence 142536, repeated endlessly with no pause between repetitions. In this sequence there is an alternation between left-triplet and right-triplet flashes. At very low speeds, there is no apparent motion at all. The lights appear simply to go on and off in sequence. At a somewhat higher speed, the true sequence (142536) is seen as a form of irregular left-and-right motion between members of the two triplets. Then, as the speed is increased, the motion appears to split into two separate streams, one involving the leftmost three lamps and the other the rightmost three. The leftmost path of motion is 1–2–3 and the rightmost one

is –4–5–6 (the dashes indicating the time periods in which the lights from the other stream are active). This segregation is exactly parallel to what happens in the auditory streaming effect. However, it is also directly explainable through Körte's third law.

This law simply states that as the speed increases, the distance between flashes must shrink if good motion is to be seen. Therefore, if we assume that potential motions between successive and nonsuccessive flashes are competing with one another for dominance, and that we finally see the one that is most dominant, the results of our example follow directly. As we speed up the sequence there is an increased tendency for shorter movements to be favored by Körte's law so that the longer between-triplet motions are suppressed in favor of the stronger within-triplet motions.

I have set up the two examples, the streaming of tones and the splitting of apparent motion, in a parallel way so that the analogy can be directly seen. Horizontal position in space is made to correspond to the frequency of the tones, with time playing the role of the second dimension in both cases.

The success of Körte's law in explaining the visual case suggests that there is a parallel law in audition, with melodic motion taking the place of spatial motion.[7] This law would state that if you want to maintain the sense of melodic motion as the frequency separation between high and low tones increases, you must slow the sequence down. As with visual apparent motion it is as if the psychological mechanism responsible for the integration of auditory sequences could not keep up with rapid changes.

Scene-Analysis Explanation

However, Körte's law is not an accident of the construction of the human brain. In both visual motion and melodic motion, the laws of grouping help to solve the scene analysis problem as the sensory input unfolds over time. In both domains, Körte's law is likely to group information appropriately. In vision it tends to group glimpses of a moving object with other glimpses of the same object rather than with those of different objects. This is important in a world where many objects can be moving at the same time and where parts of their trajectories can be hidden by closer objects such as trees. The law assumes that if a hidden object is moving a longer distance it takes it longer to get there. Hence the proportionality of distance and time that we find in the law.

The proportionality of frequency displacement and time that we observe in the streaming effect also has a value in scene analysis. What should the auditory system do if it hears a particular sound, A1, and then either a silence or an interruption by a loud sound of a different quality, and then a subsequent sound, A2, that resembles A1? Should it group A1 and A2 as coming from the same source? The auditory system assumes that the pitch of a sound tends to change continuously and therefore that the longer it has been since the sound was heard, the greater the change ought to have been. This has the effect that longer frequency jumps are tolerable only at longer time delays.

The experience of motion that we have when a succession of discrete events occurs is not a mere laboratory curiosity. When visual apparent motion is understood as a glimpse of a scene analysis process in action, new facts about it can be discovered. For example, it has been found that when the apparent

movement seems to occur in depth, in a movement slanting away from the observer, the visual system allows more time for the object to move through the third dimension than it would have if it had appeared to be moving only in the horizontal plane.[8] This happens despite the fact that although a slanting-away motion would traverse more three-dimensional space, it produces the same displacement of an object's image as a horizontal motion does on the retina of an observer. Therefore Körte's law applies to real distance in the world and not to retinal distance, and therefore can best be understood as a sophisticated part of scene analysis.

Another example of a discovery that was guided by the assumption that the rules of apparent motion exist to group glimpses of real scenes was made by Michael Mills and myself.[9] We worked with an animation sequence in which a shape disappeared from one part of a drawing and appeared in another. This change was seen as motion only if the shape was seen as representing the outline of a "figure" both before and after the disappearance. If the observer was induced to see it as "ground" (the shape of an empty space between forms) before it disappeared, and as "figure" (the shape of an actual figure) when it reappeared, the displacement was not seen as motion but as an appearance from nowhere of the figure.

Neither is the auditory streaming effect simply a laboratory curiosity. It is an oblique glimpse of a scene-analysis process doing the best it can in a situation in which the clues to the structure of the scene are very impoverished.

In general, all the Gestalt principles of grouping can be interpreted as rules for scene analysis. We can see this, for example, in the case of the principle of grouping by similarity. Consider the block-recognition problem shown earlier in figure 9.2 where the problem was to determine which areas of the drawing represented parts of the same block. Because this drawing is not very representative of the problem of scene analysis as we face it in everyday life, let us imagine it transformed into a real scene. In the natural world visible surfaces have brightness, color, and texture. It would be a good rule of thumb to prefer to group surfaces that were similar in appearance to one another on these dimensions. This would not always work, but if this principle were given a vote, along with a set of other rules of thumb, it is clear that it would contribute in a positive way to getting the right answer.

In the case of sound, the considerations are the same. If in a mixture of sounds we are able to detect moments of sound that strongly resemble one another, they should be grouped together as probably coming from the same happening. Furthermore, the closer in time two sounds that resemble each other occur, the more likely it is that they have originated with the same event. Both of these statements follow from the idea that events in the world tend to have some persistence. They do not change instantly or haphazardly. It seems likely that the auditory system, evolving as it has in such a world, has developed principles for "betting" on which parts of a sequence of sensory inputs have arisen from the same source. Such betting principles could take advantage of properties of sounds that had a reasonably high probability of indicating that the sounds had a common origin. Viewed from this perspective, the Gestalt principles are seen to be principles of scene analysis that will generally contribute to a correct decomposition of the mixture of effects that reaches our

senses. I am not claiming that the auditory system "tries" to achieve this result, only that the processes have been selected by evolution because they did achieve them.

The argument that I have made does not imply that Gestalt theory is wrong. For the Gestaltists, the phenomena of perceptual grouping arose from the fact that there were forces of attraction and segregation that operated in a perceptual field. This may indeed be the mechanism by which the grouping occurs. I am simply arguing that even if this is the form of the computation, the particular grouping force given to each property of the sensory input and the way in which the grouping forces are allowed to interact have been determined (through evolution) to be ones that will tend to contribute to the successful solution of the scene analysis problem.

Closure and Belongingness

Our senses of vision and audition, living in the same world, often face similar problems. So we should not be surprised if we often find them using similar approaches to overcome those problems. We have seen how the two systems sometimes deal with fragmented views of a sequence of events by connecting them in plausible ways. Another strong similarity between the sense modalities can be seen in the phenomenon of "perceived continuity." This is a phenomenon that is sometimes said to be an example of "perceptual closure."

The tendency to close certain "strong" perceptual forms such as circles was observed by the Gestalt psychologists. An example might be the drawing shown in figure 9.5 in which we are likely to see a circle partly obscured by an irregular form. The circle, though its outer edge is incomplete in the picture, is not seen as incomplete but as continuing on behind the other form. In other words, the circle has closed perceptually.

It is commonly said that the Gestalt principle of closure is concerned with completing forms with gaps in them. But if it did that, we would not be able to see any forms with gaps in them, which would be ridiculous. The principle is really one for completing *evidence* with gaps in it.

The Gestalt psychologists argued that closure would occur in an interrupted form if the contour was "strong" or "good" at the point of interruption. This would be true when the contours of the form continued smoothly on both sides of the interruption so that a smooth continuation could be perceived. Presumably laws of similarity would also hold so that if the regions on two sides of an interruption were the same brightness, for instance, they would be more likely to be seen as a single one continuing behind the interruption.

Like the perceptual grouping of discrete events, closure can also be seen as a scene-analysis principle. This can be illustrated with figure 9.13 which shows a number of fragments that are really parts of a familiar object or objects. The fragments were obtained by taking the familiar display and laying an irregularly shaped mask over it. Then the parts that were underneath the mask were eliminated, leaving visible only those parts that had not been covered by it.

Why do the fragments not close up perceptually in this figure? A plausible Gestalt answer might be that the forces of closure are not strong enough. The contours of the fragments might not be similar enough or in good continuation with one another. However, it is easy to show that these are not the basic rea-

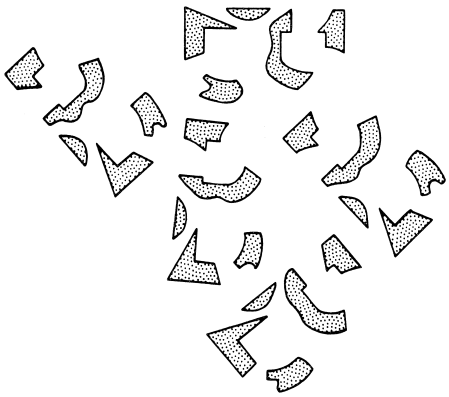


Figure 9.13
Fragments do not organize themselves strongly when there is no information for occlusion. (From Bregman 1981.)

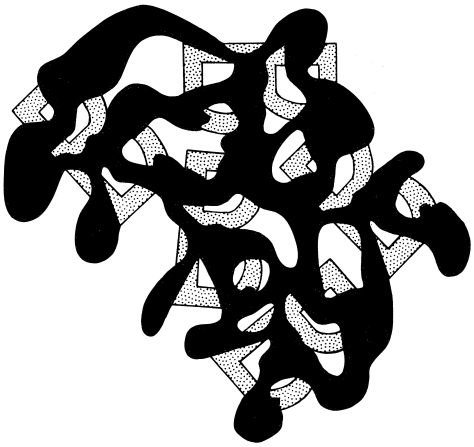


Figure 9.14
The same fragments shown in figure 9.13, except that information for occlusion has been added, causing the fragments on the boundaries of the occluding form to be grouped. (From Bregman 1981.)

sons for the lack of closure. The problem in this figure is that the visual system does not know where the evidence is incomplete. Look at what happens when the picture is shown with the mask present as in figure 9.14. The visual system quickly joins the fragments without the observer having to think about it. The Gestalt principle of closure has suddenly come alive in the presence of the mask.

What information could the mask be providing? It tells the eye two things. It explains which contours have been produced by the shape of the fragments themselves as contrasted with those that have been produced by the shape of the mask that is covering them. It also provides information about occlusion (which spaces between fragments were created by the fact that the mask

occluded our view of the underneath shape). These spaces should be ignored and treated as missing evidence, not as actual spaces. The continuity among the contours of the fragments of a particular B undoubtedly contributes to their grouping, but this continuity becomes effective only in the presence of occlusion information.

The conclusion to be reached is this: the closure mechanism is really a way of dealing with missing evidence. But before our perceptual systems are willing to employ it, they first have to be shown that some evidence is missing. This explains how we can see figures with actual gaps in them; we have no reason to believe that the missing parts are merely being hidden. Figures 9.13 and 9.14 indicate that Gestalt principles are just oblique glimpses of a process of scene analysis that looks as much like an evidence-processing system as like the simple grouping-by-attraction system described by Gestalt psychology.

There is evidence that principles of grouping act in an equally subtle way in audition. There is a problem in hearing that is much like the problem of occlusion in seeing. This is the phenomenon of masking. Masking occurs when a loud sound covers up or drowns out a softer one. Despite the masking, if the softer sound is longer, and can be heard both before and after a brief burst of the louder one, it can be heard to continue behind the louder one just as B's were seen as continuing behind the occluding blob in figure 9.14, and as the circle seemed to continue behind the occluding form in the example of figure 9.5. What is more, even if the softer sound is *physically removed* during the brief loud sound, it is still heard as continuing through the interruption.

This illusion has many names, but I will refer to it as the illusion of continuity. It occurs with a wide range of sounds. An example is shown in figure 9.15 where an alternately rising and falling pure-tone glide is periodically interrupted by a short loud burst of broad-band noise (like the noise between stations on a radio). When the glide is broken at certain places but no masking sound is present during the breaks, as in the left panel, the ear hears a series of rising and falling glides, but does not put them together as a single sound any more than the eye puts together the fragments of figure 9.13. However, if the masking noise is introduced in the gaps so as to exactly cover the silent spaces, as in the right panel, the ear hears the glide as one continuous rising and falling sound passing right through the interrupting noise. The integration of the continuous glide pattern resembles the mental synthesis of B's in figure 9.14. They are both effortless and automatic.

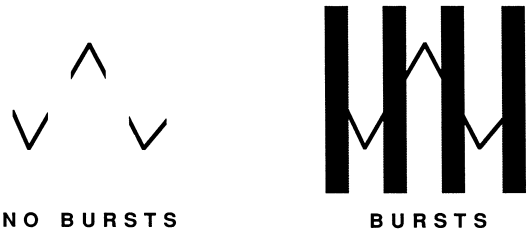


Figure 9.15
Tonal glides of the type used by Dannenbring (1976). Left: the stimulus with gaps. Right: the stimulus when the gaps are filled with noise.

Again you could see the auditory effect as an example of the Gestalt principle of closure. However another way of looking at it may be more profitable. Richard Warren has interpreted it as resulting from an auditory mechanism that compensates for masking.[10] He has shown that the illusion can be obtained only when the interrupting noise would have masked the signal if it had really been there. The interrupting noise must be loud enough and have the right frequency components to do so. Putting that in the context of this chapter, we see that the illusion is another oblique glance of the auditory scene-analysis process in action.

We have seen how two types of explanation, one deriving from Gestalt psychology and the other derived from considerations of scene analysis, have been applicable to both the streaming and continuity effects. They differ in style. The Gestalt explanation sees the principles of grouping as phenomena in themselves, a self-sufficient system whose business it is to organize things. The scene-analysis approach relates the process more to the environment, or, more particularly, to the problem that the environment poses to the perceiver as he or she (or it) tries to build descriptions of environmental situations.

Sequential versus Spectral Organization

Perceptual Decomposition of Complex Sounds

We have looked at two laboratory phenomena in audition that show the activity of the scene-analysis process: the streaming effect and the illusory continuation of one sound behind another. There is a third phenomenon that deserves to be mentioned in this introductory chapter. It is introduced here not to demonstrate a parallel between vision and audition, but to show another dimension of the grouping problem. This is the perceptual decomposition of simultaneous sounds. It can be illustrated through an experiment by Bregman and Pinker.[11]

The sounds used in this experiment are shown in figure 9.16. They consist of a repeating cycle formed by a pure tone A alternating with a complex tone that has two pure-tone components, B and C. This is inherently an ambiguous event. For example, it could be created by giving an audio oscillator to each of two people. The oscillator given to one of them puts out the pure tone A, while the one given to the other puts out the complex tone BC. The two persons are

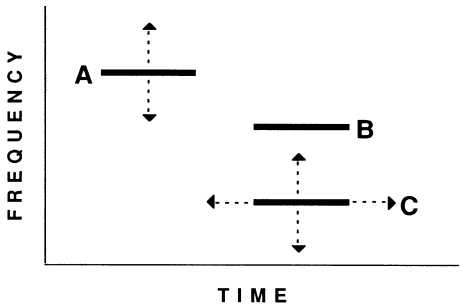


Figure 9.16
Stimulus used by Bregman and Pinker (1978). A, B, and C are pure tone components.

asked to play their oscillators in rapid alternation. If this were the way the sound had been created, the correct perceptual analysis would be to hear a pure tone alternating with a rich-sounding complex tone. This, however, is only one possibility for the origin of the sound. The second is that we have given out oscillators, as before, to two persons. This time, however, both of the oscillators can put out only pure tones. One person is told to sound his instrument twice on each cycle to make the tones A and B, whereas the other is told to play his tone only once on each cycle to make the tone C. He is told to synchronize his C tone with the B tone of his partner. If our auditory systems were to correctly represent the true causes of the sound in this second case, we should hear two streams: one consisting of the repetitions of tones A and B, accompanied by a second that contains only the repetitions of tone C. In this way of hearing the sequence, there should be no rich tone BC because the richness is an accidental by-product of the mixture of two signals. If the auditory system is built to hear the properties of meaningful events rather than of the accidental by-products of mixtures, it should discard the latter.

The experiment showed that it was possible to hear the sequence in either way, depending on two factors. The first was the frequency proximity of tones A and B. The closer they were to one another in frequency, the greater the likelihood of hearing A and B as forming a single stream separate from C. Apparently the auditory system uses the proximity of a succession of frequencies, much as it does in the case of the streaming phenomenon, as evidence that they are from a common source. The second factor was the synchrony of tones B and C. If their onsets and offsets were synchronized, they tended to be fused and heard as a single complex sound BC, which was heard as alternating with A. Furthermore, the effects of the BC synchrony were competitive with the effects of the AB frequency proximity. It was as if A and C were competing to see which one would get to group with C. If the synchrony of C with B was reduced, B would be more likely to group with A, unless, of course, the AB connection was made weaker by moving A further away in frequency from B.

Horizontal and Vertical Processes of Organization

There is a distinction that ought to be made now because it follows directly from the Bregman-Pinker experiment. This is the distinction between the processes of sequential and spectral integration.

The process of putting A and B together into a stream can be referred to as sequential integration. This is the kind of integration that forms the melodic component of music. It is the process that connects events that have arisen at different times from the same source. It uses the changes in the spectrum and the speed of such changes as major clues to the correct grouping. The sequential process is what is involved in the streaming effect that was discussed earlier.

The fusing of B with C into a single sound is what will be referred to as simultaneous integration or, in special contexts, as spectral integration, a term borrowed from James Cutting.[12] It is this process that takes acoustic inputs that occur at the same time, but at different places in the spectrum or in space, and treats them as properties of a single sound. It is responsible for the fact that we can interpret a single spectrum of sound as arising from the mixture of two or more sound sources, with the timbre of each one being computed from just

those spectral components that have been allocated to that source. This happens, for example, when we hear two singers, one singing "ee" and the other "ah," on different pitches. Despite the fact that all we have is a single spectrum, with the harmonics from the two voices intermixed, we can clearly hear the two vowels. Since a vowel sound is a sort of timbre, this example shows that we can extract two timbres at the same time from a single signal.

If we turn back to the mixed spectrogram shown in figure 9.4, we see that in order to put together the streaks of darkness belonging to the same acoustic source, the same two kinds of grouping are necessary: (1) putting together events that follow one another in time (sequential grouping) and (2) integrating components that occur at the same time in different parts of the spectrum (simultaneous grouping). Musicians speak of a horizontal and a vertical dimension in written music. By horizontal, they refer to the groupings across the page that are seen as melody. By vertical, they refer to the simultaneous events that form chords and harmony. These are the same two dimensions as the ones called sequential and simultaneous.

It is useful to distinguish these two aspects of organization because they are controlled by different acoustic factors. Of course they interact, too.

Types of Explanation of These Phenomena

It is interesting to take a moment to see how these phenomena are related to various theoretical positions. I will consider their relation to concepts drawn from computer modeling, syntactic theory, Gestalt psychology, and physiological explanation.

The computer modeling approach has contributed an important idea: the notion of a heuristic. The idea was evolved in the process of designing computer programs to solve difficult problems for which no mathematical solution was known. The approach taken by the designers was to employ heuristics, which are defined as procedures that are not guaranteed to solve the problem, but are likely to lead to a good solution. An example would be the use of heuristic tests by computer chess programs to determine whether a proposed move would lead to a good position (e.g., to test whether the move would result in the computer controlling the center of the board or whether the move would lead to an exchange of pieces that favored the computer). Each move is evaluated by a number of such heuristics. No one of them can guarantee success, but if there are a large number, each with some basis in the structure of the game of chess, a move that satisfies most of them will probably be a good one. Furthermore, if each of the heuristic evaluation processes has a chance to vote for or against the move, the program will be less likely to be tricked than it would be if it based its move on only one or two criteria, no matter how good they were.

I believe that the perceptual systems work in similar ways. Having evolved in a world of mixtures, humans have developed heuristic mechanisms capable of decomposing them. Because the conditions under which decomposition must be done are extremely variable, no single method is guaranteed to succeed. Therefore a number of heuristic criteria must be used to decide how to group the acoustic evidence. These criteria are allowed to combine their effects in a process very much like voting. No one factor will necessarily vote correctly,

but if there are many of them, competing with or reinforcing one another, the right description of the input should generally emerge. If they all vote in the same way, the resulting percept is stable and unambiguous. When they are faced with artificial signals, set up in the laboratory, in which one heuristic is made to vote for integration and another for segregation, the resulting experiences can be unstable and ambiguous.

My use of the word "heuristic" does not imply a computer-like procedure that involves a long sequence of steps, extended over time. We have to bear in mind that the decisions of the auditory system are carried out in very short periods of time. I use the word heuristic in its functional sense only, as a process that contributes to the solution of a problem.

Whereas the perceptual phenomena that we examined earlier are the province of psychologists, the problem of how people build mental descriptions is a topic that has been looked at by linguists too. As a result, they have provided us with a metaphor for understanding auditory scene analysis. This metaphor, "deep structure," derives from the study of the syntactic structure of sentences.

One of the basic problems in syntax is how to describe the rules that allow the speaker to impose a meaning on a sentence by adding, subtracting, or rearranging elements in the sentence. For example, in English one of these rules imposes the form of a question on a sentence by placing the auxiliary verb at the beginning of the sentence. Thus, the active sentence "He has gone there" is expressed in a question as "Has he gone there?" The difficulty that occurs when a language loads a sentence with meanings is that when a large number of form-shaping rules are piled on top of one another, it becomes difficult to untangle them and to appreciate the contribution of each of them to the final product. Somehow all speakers of English come to be able to do this, but the learning takes some time. In the 1960s, Noam Chomsky introduced the notion of the "deep structure" of a sentence, a description of a sentence that separately and explicitly described all the underlying syntactic forms and displayed their interrelationships. When a theorist, or a listener, starts with a given sentence and builds a description of its syntax, this is called "parsing" the sentence. It was argued by psychologists who were inspired by Chomsky's approach that in the course of understanding a sentence, the hearer parses a sentence and builds a deep structure for it.

We can talk about perception in a very similar way. Just as a spoken sentence imposes an extraordinary decoding problem upon the listener, so does a nonlinguistic sensory input. Whenever we experience an event, the sensory impression is always the result of an elaborate composition of physical influences. If we look at a four-inch-square area of a table top, for example, the local properties of this area have been affected by many factors: the table's shininess, the variations in its surface color, the unevenness of its surface, the shadow of a nearby object, the color of the light source, the slant of the surface of the table relative to our eyes, and perhaps many more. These factors are all simultaneously *shaping* the sensory information; they are not simply inserted side by side. The shininess is not at one place in our visual image, the surface color at another, and so on. Neither can they be extracted from the sense data independently of one another.

The same thing happens in audition. If we look at any one-tenth-second slice of figure 9.4, the information shown in that slice represents a composition of influences. The spectrum may have been shaped by voices and by other simultaneous sounds. Somehow, if we are able to understand the events that have shaped it, we are succeeding, as in sentence comprehension, in developing a mental description that displays the simple causative factors and their interrelationships in an explicit way.

There is a provocative similarity among the three examples—the syntactical, the visual, and the auditory. In all three cases, the perceivers are faced with a complex *shaping* of the sensory input by the effects of various simple features, and they must recover those features from their effects. Transposing the linguist's vocabulary to the field of perception, one might say that the job of the perceiver is to parse the sensory input and arrive at its deep structure. In some sense the perceiver has to build up a description of the regularities in the world that have shaped the evidence of our senses. Such regularities would include the fact that there are solid objects with their own shapes and colors (in vision) and sounds with their own timbres and pitches (in audition).

Although the approach of this chapter is not physiological, it is important to see its relation to physiological explanation. We can take as an example the physiological explanations that have been offered for the streaming effect of figure 9.8. It has been proposed that the segregation into two streams occurs because a neural mechanism responsible for tracking changes in pitch has temporarily become less effective.[13] This interpretation is supported by the results of experiments that show that the segregation becomes stronger with longer repetitions of the cycle of tones. Presumably the detector for change has become habituated in the same manner as other feature detectors are thought to. This view of the stream segregation phenomenon sees it as a breakdown. This seems to be in serious conflict with the scene-analysis view presented earlier, in which stream segregation was seen as an accomplishment. So which is it to be, breakdown or accomplishment?

We do not know whether or not this physiological explanation is correct. But even if it is, its truth may not affect the scene analysis explanation of streaming. To demonstrate why, it is necessary to again appeal to an argument based on evolution. Every physiological mechanism that develops must stand the test of the winnowing process imposed by natural selection. However, the survival of an individual mechanism will often depend not just on what it does in isolation, but on the success of the larger functional system of which it forms a part.

Because of the indirect way in which the individual physiological mechanism contributes to the successful accomplishments displayed by the larger system, it is possible that what looks like a breakdown when seen at the single-mechanism level is actually contributing to an accomplishment at the system level. To take a homespun example, consider the case of a pitfall trap. When the top of the trap, covered with branches and leaves, "breaks down" and the animal falls through into the hole, we can see that the physical breakdown (of the trap cover) represents a functional success (of the entrapment). The breakdown and the achievement are at different levels of abstraction. By analogy, it would not be contradictory to assert that the streaming effect represented both

the breakdown of a physiological mechanism and the accomplishment of scene analysis. This example illustrates how indirect the relation can be between function and physiology.

Scene-Analysis View Prevents Missing of Vision-Audition Differences

It was argued in the earlier discussion that Gestalt explanations had to be supplemented by ones based on scene analysis because the latter might lead us to new phenomena, such as the role of the occluding mask in perceptual closure. There is another difference between the two approaches. Because the Gestalt theorists saw the principles of organization as following from general properties of neural tissue they focused on similarities between the senses rather than on differences. The laws of grouping were stated in a general way, in terms of adjectives (such as "proximity" or "similarity") that could apply equally well to different sense modalities. This has had both useful and harmful effects. On the positive side it has promoted the discovery of the similar way in which perceptual organization works in different sense modalities. For example, the similarities between apparent movement and auditory streaming have become apparent. However, an exclusive focus on the common Gestalt principles, neglecting the unique scene-analysis problems that each sense must solve, is likely to neglect differences between them and cause us to miss some excellent opportunities to study special problems in audition that make themselves evident once we consider the dissimilarities between the senses. The way to get at them is to consider the differences in the way in which information about the properties of the world that we care about are carried in sound and in light. The fact that certain Gestalt principles actually are shared between the senses could be thought of as existing because they are appropriate methods for scene analysis in both domains.

As an example of the way that the scene-analysis approach can reveal important differences between the senses, let us go through the exercise of considering the roles of direct energy, reflected energy, and their mixture in the two senses.

Differences in the Ecology of Vision and Audition

There is a crucial difference in the way that humans use acoustic and light energy to obtain information about the world. This has to do with the dissimilarities in the ecology of light and sound. In audition humans, unlike their relatives the bats, make use primarily of the sound-emitting rather than the sound-reflecting properties of things. They use their eyes to determine the shape and size of a car on the road by the way in which its surfaces reflect the light of the sun, but use their ears to determine the intensity of the crash by receiving the energy that is emitted when this event occurs. The shape reflects energy; the crash creates it. For humans, sound serves to supplement vision by supplying information about the nature of events, defining the "energetics" of a situation.

There is another difference that is very much related to this one: sounds go around corners. Low-frequency sound bends around an obstruction while higher frequency sound bounces around it. This makes it possible for us to

have a distant early warning system. The reader might be tempted to object that light too goes around corners. Although it does not bend around, in the way that low-frequency sound does, it often gets around by reflection; in effect, it bounces around the corner. But notice what a difference this bouncing makes in how we can use the light. Although the bounced-around light provides illumination that allows us to see the shapes of things on our own side of the corner, unless it has been bounced by means of mirrors it has lost the shape information that it picked up when it reflected off the objects on the opposite side. Sound is used differently. We use it to discover the time and frequency pattern of the source, not its spatial shape, and much of this information is retained even when it bends or bounces around the corner.

This way of using sound has the effect, however, of making acoustic events transparent; they do not occlude energy from what lies behind them. The auditory world is like the visual world would be if all objects were very, very transparent and glowed in sputters and starts by their own light, as well as reflecting the light of their neighbors. This would be a hard world for the visual system to deal with.

It is not true then that our auditory system is somehow more primitive simply because it does not deliver as detailed information about the shapes, sizes, and surface characteristics of objects. It simply has evolved a different function and lives in a different kind of world.

What of echoes? We never discuss echoes in light because its speed is so fast and the distances in a typical scene are so small that the echo arrives in synchrony with the original signal. Furthermore, in vision we are usually interested in the echoes, not the original signal, and certainly not in integrating the two into a single image. Light bounces around, reflecting off many objects in our environments, and eventually gets to our eyes with the imprint of the unoccluded objects still contained in it. Because the lens-and-retina system of the eye keeps this information in the same spatial order, it allows us access to the information about each form separately. Echoes are therefore very useful in specifying the shapes of objects in vision because the echoes that come off different surfaces do not get mixed together on the way to our eye.

The case is otherwise in audition. Because our ears lack the lenses that could capture the spatial layout of the echoes from different surfaces, we are usually interested in the source of sound rather than in the shapes of objects that have reflected or absorbed it. The individual spatial origins of the parts of a reflected wave front are barely preserved at all for our ears. Therefore, when the sound bounces off other objects and these echoes mix with the original signal, they obscure the original properties of the sound. Although echoes are delayed copies and, as such, contain all the original structure of the sound, the mixing of the original and the echo creates problems in using this redundant structural information effectively.

The two senses also make different uses of the absorption of energy by the environment. The fact that different objects absorb light in different ways gives them their characteristic colors and brightnesses, but this differential absorption is not as valuable in hearing because our ears cannot separate the reflections from small individual objects. We do hear the "hardness" or "softness" of the entire room that we are in. This corresponds to the color information carried in

light, but the acoustic information is about very large objects, whereas the information in light can be about very small ones.

In summary, we can see that the differences in how we use light and sound create different opportunities and difficulties for the two perceptual systems and that they probably have evolved specialized methods for dealing with them.

Primitive versus Schema-Based Stream Segregation

It seems reasonable to believe that the process of auditory scene analysis must be governed by both innate and learned constraints. The effects of the unlearned constraints are called "primitive segregation" and those of the learned ones are called "schema-based segregation."

One reason for wanting to think that there are unlearned influences on segregation is the fact that there are certain constant properties of the environment that would have to be dealt with by every human everywhere. Different humans may face different languages, musics, and birds and animals that have their own particular cries. A desert certainly sounds different from a tropical forest. But certain essential physical facts remain constant. When a harmonically structured sound changes over time, all the harmonics in it will tend to change together in frequency, in amplitude, and in direction, and to maintain a harmonic relationship. This is not true of just some particular environment but of broad classes of sounds in the world.

Such regularities can be used in reverse to infer the probable underlying structure of a mixture. When frequency components continue to maintain a harmonic relationship to one another despite changes in frequency, amplitude, and spatial origin, they will almost always have been caused by a coherent physical event. The later chapters show that the human auditory system makes use of such regularity in the sensory input. But is this innate? I think that it is. The internal organs of animals evolve to fit the requirements of certain constant factors in their environments. Why should their auditory systems not do likewise?

Roger Shepard has argued for a principle of "psychophysical complementarity," which states that the mental processes of animals have evolved to be complementary with the structure of the surrounding world.[14] For example, because the physical world allows an object to be rotated without changing its shape, the mind must have mechanisms for rotating its representations of objects without changing their shapes. The processes of auditory perception would fall under this principle of complementarity, the rules of auditory grouping being complementary with the redundancies that link the acoustic components that have arisen from the same source.

The Gestalt psychologists argued that the laws of perceptual organization were innate. They used two types of evidence to support their claim. One was the fact that the phenomenon of camouflage, which works by tricking the organizational processes into grouping parts of an object with parts of its surroundings, could be made to disguise even highly familiar shapes. Clearly, then, some general grouping rules were overriding learned knowledge about

the shape of objects. The second was the fact that perceptual organization could be demonstrated with very young animals.

To the arguments offered by the Gestaltists can be added the following one: From an engineering point of view, it is generally easier to design a machine that can do some task directly than to design one that can *learn* to do it. We can design machines that can parse or generate fairly complex sentences, but there has been limited success in designing one that could learn grammatical rules from examples without any designed-in knowledge of the formal structure of those rules. By analogy, if you think of the physical world as having a "grammar" (the physical laws that are responsible for the sensory impressions that we receive), then each human must be equipped either with mechanisms capable of learning about many of these laws from examples or with a mechanism whose genetic program has been developed once and for all by the species as a result of billions of parallel experiments over the course of history, where the lives of the members of the species and its ancestors represent the successes and the lives of countless extinct families the failures. To me, evolution seems more plausible than learning as a mechanism for acquiring at least a general capability to segregate sounds. Additional learning-based mechanisms could then refine the ability of the perceiver in more specific environments.

The innate influences on segregation should not be seen as being in opposition to principles of learning. The two must collaborate, the innate influences acting to "bootstrap" the learning process. In language, meaning is carried by words. Therefore if a child is to come to respond appropriately to utterances, it is necessary that the string be responded to in terms of the individual words that compose it. This is sometimes called the segmentation problem. Until you look at a spectrogram of continuous speech occurring in natural utterances, the task seems easy. However, on seeing the spectrogram, it becomes clear that the spaces that we insert into writing to mark the boundaries of words simply do not occur in speech. Even if sentences were written without spaces, adults could take advantage of prior knowledge to find the word boundaries. Because they already know the sequences of letters that make meaningful words, they could detect each such sequence and place tentative word boundaries on either side of it. But when infants respond to speech they have no such prior learning to fall back on. They would be able to make use only of innate constraints. I suspect a main factor used by infants to segment their first words is acoustic discontinuity. The baby may hear a word as a unit only when it is presented in isolation, that is, with silence (or much softer sound) both before and after it. This would be the result of an innate principle of boundary formation. If it were presented differently, for example, as part of a constant phrase, then the phrase and not the word would be treated as the unit. The acoustic continuity within a sample of speech and the discontinuities at its onset and termination would be available, even at the earliest stage of language acquisition, to label it as a single whole when it was heard in isolation. Once perceived as a whole, however, its properties could be learned. Then, after a few words were learned, recognition mechanisms could begin to help the segmentation process. The infant would now be able to use the beginnings and ends of these familiar patterns to establish boundaries for other words that might lie between them. We

can see in this example how an innate grouping rule could help a learning process to get started. (I am not suggesting that the establishing of acoustic boundaries at discontinuities is the only method that infants use to discover units, but I would be very surprised if it were not one of them.)

Another example of innate segregation that was given earlier concerned an infant trying to imitate an utterance by her mother. It was argued that the fact that the infant did not insert into her imitation the cradle's squeak that had occurred during her mother's speech displayed her capacity for auditory scene analysis. I am also proposing that this particular capacity is based on innately given constraints on organization.

There is much experimental evidence drawn from experiments on the vision of infants that supports the existence of innate constraints on perceptual organization. Corresponding experiments on auditory organization, however, are still in short supply.

One such study was carried out by Laurent Demany in Paris.[15] Young infants from $1\frac{1}{2}$ to $3\frac{1}{2}$ months of age were tested with sequences of tones. The method of habituation and dishabituation was used. This is a method that can be used with infants to discover whether they consider two types of auditory signals the same or different. At the beginning, a sound is played to the babies every time they look at a white spot on a screen in front of them. The sound acts as a reward and the babies repeatedly look at the white spot to get the interesting sound. After a number of repetitions of this "look and get rewarded" sequence, the novelty of the sound wears off and it loses its potency as a reward (the infants are said to have habituated to the sound). At this point the experimenter replaces the sound by a different one. If the newness of the sound restores its ability to act as a reward, we can conclude that the infants must consider it to be a different sound (in the language of the laboratory, they have become dishabituated), but if they continue ignoring it, they must consider it to be the same as the old one.

Using this method, Demany tried to discover whether infants would perceptually segregate high tones from low ones. The proof that they did so was indirect. The reasoning went as follows: Suppose that four tones, all with different pitches, are presented in a repeating cycle. Two are higher in pitch (H1 and H2) and two are lower (L1 and L2), and they are presented in the order H1, L1, H2, L2, ... If the high and low tones are segregated into different perceptual streams, the high stream will be heard as

H1–H2–H1–H2–H1–H2–. . .

and the low stream will be perceived as

L1–L2–L1–L2–L1–L2–. . .

(where the dashes represent brief within-stream silences). In each stream all that is heard is a pair of alternating tones.

Now consider what happens when the reverse order of tones is played, namely L2, H2, L1, H1, ... If the high tones segregate from the low ones, the high stream is heard as

H2–H1–H2–H1–H2–H1–. . .

and the low one as

L2–L1–L2–L1–L2–L1–. . . .

Again each stream is composed of two alternating tones. In fact, if the infant lost track of which one of the pair of tones started the sequence, the two streams would be considered to be exactly the same as they were with the original order of tones. Suppose, however, that the infant does not segregate the high from the low tones. In this case the forward and the backward orders of tones are quite different from one another and remain so even if the infant forgets which tone started the sequence.

To summarize, the segregated streams are quite similar for the forward and backward sequences whereas the unsegregated sequences are quite different. Using the habituation/dishabituation method, Demany tried to determine whether the infants considered the forward and backward sequences the same or different. The results showed that they were reacted to as being the same. This implied that stream segregation had occurred. In addition, Demany showed that this result was not due to the fact that the infants were incapable in general of distinguishing the order of tonal sequences. Pairs of sequences whose segregated substreams did not sound similar to an adult were not reacted to as being the same by infants. In general, the infant results paralleled those of adult perception and the older and younger infants did not differ in their reactions.

Undoubtedly more such research is required. After all, the infants were not newborns; they had had some weeks of exposure to the world of sound. But after this pioneering study, the burden of proof shifts to those who would argue that the basic patterns of auditory organization are learned. Unfortunately, working with very young infants is difficult and the amount of data collected per experiment is small.

The unlearned constraints on organization can clearly not be the only ones. We know that a trained musician, for example, can hear the component sounds in a mixture that is impenetrable to the rest of us. I have also noticed that when researchers in my laboratory prepare studies on perceptual organization, they must listen to their own stimuli repeatedly. Gradually their intuitions about how easy it is to hear the stimulus in a particular way comes to be less and less like the performance of the untrained listeners who are to serve as the subjects of the experiment.

Undoubtedly there are learned rules that affect the perceptual organization of sound. I shall refer to the effects of these rules as "schema-based integration" (a schema is a mental representation of some regularity in our experience). Schema-based analysis probably involves the learned control of attention and is very powerful indeed. The learning is based on the encounter of individuals with certain lawful patterns of their environments, speech and music being but two examples. Since different environments contain different languages, musics, speakers, animals, and so on, the schema-based stream segregation skills of different individuals will come to have strong differences, although they may have certain things in common. In later chapters, I will give some examples of the effects of schema-governed scene analysis in the fields of music and language, and will discuss a theory of sequential integration of sound, pro-

posed by Mari Reiss Jones, that is best understood as describing the influence of schemas on stream segregation.

Verification of the Theory

The theory presented in this chapter proposes that there is an auditory stream-forming process that is responsible for a number of phenomena such as the streaming effect and the illusion of continuity, as well as for the everyday problems of grouping components correctly to hear that a car is approaching as we cross a street, or "hearing out" a voice or an instrument from a musical performance. This is not the type of theory that is likely to be accepted or rejected on the basis of one crucial experiment. Crucial experiments are rare in psychology in general. This is because the behavior that we observe in any psychological experiment is always the result of a large number of causal factors and is therefore interpretable in more than one way. When listeners participate in an experiment on stream segregation, they do not merely perceive; they must remember, choose, judge, and so on. Each experimental result is always affected by factors outside the theory, such as memory, attention, learning, and strategies for choosing one's answer. The theory must therefore be combined with extra assumptions to explain any particular outcome. Therefore it cannot easily be proven or falsified.

Theories of the type I am proposing do not perform their service by predicting the exact numerical values in experimental data. Rather they serve the role of guiding us among the infinite set of experiments that could be done and relationships between variables that could be studied. The notion of stream segregation serves to link a number of causes with a number of effects. Stream segregation is affected by the speed of the sequence, the frequency separation of sounds, the pitch separation of sounds, the spatial location of the sounds, and many other factors. In turn, the perceptual organization into separate streams influences a number of measurable effects, such as the ability to decide on the order of events, the tendency to hear rhythmic factors within each segregated stream, and the inability to judge the order of events that are in different streams. Without the simplifying idea of a stream-forming process, we would be left with a large number of empirical relations between individual causal influences and measurable behaviors.

A theory of this type is substantiated by converging operations. This means that the concepts of "perceptual stream" and "scene-analysis process" will gain in plausibility if a large number of different kinds of experimental tasks yield results that are consistent with these ideas.

Summary

I started this chapter with a general introduction to a number of problems. I began with the claim that audition, no less than vision, must solve very complex problems in the interpretation of the incoming sensory stimulation. A central problem faced by audition was in dealing with mixtures of sounds. The sensory components that arise from distinct environmental events have to be segregated into separate perceptual representations. These representations

(which I called streams) provide centers of description that connect sensory features so that the right combinations can serve as the basis for recognizing the environmental events. This was illustrated with three auditory phenomena, the streaming effect, the decomposition of complex tones (the ABC experiment), and perceptual closure through occluding sounds.

The explanation that I offered had two sides. It discussed both perceptual representations and the properties of the acoustic input that were used heuristically to do the segregation. I argued that one had to take the ecology of the world of sound into account in looking for the methods that the auditory system might be using, and claimed that this could serve as a powerful supplement to the Gestalt theorist's strategy of looking for formal similarities in the activity of different senses. Finally I proposed that there were two kinds of constraints on the formation of perceptual representations, unlearned primitive ones and more sophisticated ones that existed in learned packages called schemas.

Notes

1. For example, those described by Deutsch (1975a).
2. From Guzman (1969).
3. Bregman and Rudnicky (1975).
4. Treisman and Schmidt (1982).
5. Julesz and Hirsh (1972).
6. Forms of this effect have been described by Vicario (1965, 1982) and Bregman and Achim (1973).
7. See discussion in van Noorden (1975). A more elaborate form of Körte's law in audition has been offered by Jones (1976).
8. Ogasawara (1936), Corbin (1942), and Attneave and Block (1973).
9. Bregman and Mills (1982).
10. See review in Warren (1982).
11. Bregman and Pinker (1978).
12. Cutting (1976).
13. Anstis and Saida (1985).
14. Shepard (1981).
15. Demany (1982).

References

Anstis, S., and Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 257–271.

Attneave, F., and Block, G. (1973). Apparent movement in tridimensional space. *Perception & Psychophysics, 13*, 301–307.

Bregman, A. S. (1981). Asking the "what for" question in auditory perception. In M. Kubovy and J. R. Pomerantz (eds.), *Perceptual Organization*. Hillsdale, N.J.: Erlbaum.

Bregman, A. S., and Achim, A. (1973). Visual stream segregation. *Perception & Psychophysics, 13*, 451–454.

Bregman, A. S., and Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89*, 244–249.

Bregman, A. S., and Mills, M. I. (1982). Perceived movement: The Flintstone constraint. *Perception, 11*, 201–206.

Bregman, A. S., and Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology, 32*, 19–31.

Bregman, A. S., and Rudnicky, A. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance, 1*, 263–267.

Corbin, H. H. (1942). The perception of grouping and apparent movement in visual depth. *Archives of Psychology*, no. 273.

Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review, 83*, 114–140.

Dannenbring, G. L. (1976). Perceived auditory continuity with alternately rising and falling frequency transitions. *Canadian Journal of Psychology, 30*, 99–114.

Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behavior and Development, 5*, 261–276.

Deutsch, D. (1975). Musical illusions. *Scientific American, 233*, 92–104.

Guzman, A. (1969). Decomposition of a visual scene into three-dimensional bodies. In A. Grasselli (ed.), *Automatic Interpretation and Classification of Images*. New York: Academic Press.

Jones, M. R. (1976). Time, our lost dimensions: Toward a new theory of perception, attention, and memory. *Psychology Review, 83*, 323–355.

Julesz, B., and Hirsh, I. J. (1972). Visual and auditory perception—An essay of comparison. In E. E. David, Jr., and P. B. Denes (eds.), *Human Communication: A Unified View*. New York: McGraw-Hill.

Ogasawara, J. (1936). Effect of apparent separation on apparent movement. *Japanese Journal of Psychology, 11*, 109–122.

PART VI

Categories and Concepts

Chapter 10

Principles of Categorization

Eleanor Rosch

The following is a taxonomy of the animal kingdom. It has been attributed to an ancient Chinese encyclopedia entitled the *Celestial Emporium of Benevolent Knowledge*:

> On those remote pages it is written that animals are divided into (a) those that belong to the Emperor, (b) embalmed ones, (c) those that are trained, (d) suckling pigs, (e) mermaids, (f) fabulous ones, (g) stray dogs, (h) those that are included in this classification, (i) those that tremble as if they were mad, (j) innumerable ones, (k) those drawn with a very fine camel's hair brush, (l) others, (m) those that have just broken a flower vase, (n) those that resemble flies from a distance. (Borges 1966, p. 108)

Conceptually, the most interesting aspect of this classification system is that it does not exist. Certain types of categorizations may appear in the imagination of poets, but they are never found in the practical or linguistic classes of organisms or of man-made objects used by any of the cultures of the world. For some years, I have argued that human categorization should not be considered the arbitrary product of historical accident or of whimsy but rather the result of psychological principles of categorization, which are subject to investigation. This chapter is a summary and discussion of those principles.

The chapter is divided into five parts. The first part presents the two general principles that are proposed to underlie categorization systems. The second part shows the way in which these principles appear to result in a basic and primary level of categorization in the levels of abstraction in a taxonomy. It is essentially a summary of the research already reported on basic level objects (Rosch et al., 1976). Thus the second section may be omitted by the reader already sufficiently familiar with that material. The third part relates the principles of categorization to the formation of prototypes in those categories that are at the same level of abstraction in a taxonomy. In particular, this section attempts to clarify the operational concept of prototypicality and to separate that concept from claims concerning the role of prototypes in cognitive processing, representation, and learning for which there is little evidence. The fourth part presents two issues that are problematical for the abstract principles of categorization stated in the first part: (1) the relation of context to basic level objects and prototypes; and (2) assumptions about the nature of the attributes of real-world objects that underlie the claim that there is structure in the world.

From chapter 8 in *Concepts: Core Readings*, ed. E. Margolis and S. Laurence (Cambridge, MA: MIT Press, 1978/1999), 189–206. Reprinted with permission.

The fifth part is a report of initial attempts to base an analysis of the attributes, functions, and contexts of objects on a consideration of objects as props in culturally defined events.

It should be noted that the issues in categorization with which we are primarily concerned have to do with explaining the categories found in a culture and coded by the language of that culture at a particular point in time. When we speak of the formation of categories, we mean their formation in the culture. This point is often misunderstood. The principles of categorization proposed are not as such intended to constitute a theory of the development of categories in children born into a culture nor to constitute a model of how categories are processed (how categorizations are made) in the minds of adult speakers of a language.

The Principles

Two general and basic principles are proposed for the formation of categories: The first has to do with the function of category systems and asserts that the task of category systems is to provide maximum information with the least cognitive effort; the second has to do with the structure of the information so provided and asserts that the perceived world comes as structured information rather than as arbitrary or unpredictable attributes. Thus maximum information with least cognitive effort is achieved if categories map the perceived world structure as closely as possible. This condition can be achieved either by the mapping of categories to given attribute structures or by the definition or redefinition of attributes to render a given set of categories appropriately structured. These principles are elaborated in the following.

Cognitive Economy

The first principle contains the almost common-sense notion that, as an organism, what one wishes to gain from one's categories is a great deal of information about the environment while conserving finite resources as much as possible. To categorize a stimulus means to consider it, for purposes of that categorization, not only equivalent to other stimuli in the same category but also different from stimuli not in that category. On the one hand, it would appear to the organism's advantage to have as many properties as possible predictable from knowing any one property, a principle that would lead to formation of large numbers of categories with as fine discriminations between categories as possible. On the other hand, one purpose of categorization is to reduce the infinite differences among stimuli to behaviorally and cognitively usable proportions. It is to the organism's advantage not to differentiate one stimulus from others when that differentiation is irrelevant to the purposes at hand.

Perceived World Structure

The second principle of categorization asserts that unlike the sets of stimuli used in traditional laboratory-concept attainment tasks, the perceived world is not an unstructured total set of equiprobable co-occurring attributes. Rather, the material objects of the world are perceived to possess (in Garner's, 1974,

sense) high correlational structure. That is, given a knower who perceives the complex attributes of feathers, fur, and wings, it is an empirical fact provided by the perceived world that wings co-occur with feathers more than with fur. And given an actor with the motor programs for sitting, it is a fact of the perceived world that objects with the perceptual attributes of chairs are more likely to have functional sit-on-able-ness than objects with the appearance of cats. In short, combinations of what we perceive as the attributes of real objects do not occur uniformly. Some pairs, triples, etc., are quite probable, appearing in combination sometimes with one, sometimes another attribute; others are rare; others logically cannot or empirically do not occur.

It should be emphasized that we are talking about the perceived world and not a metaphysical world without a knower. What kinds of attributes *can* be perceived are, of course, species-specific. A dog's sense of smell is more highly differentiated than a human's, and the structure of the world for a dog must surely include attributes of smell that we, as a species, are incapable of perceiving. Furthermore, because a dog's body is constructed differently from a human's, its motor interactions with objects are necessarily differently structured. The "out there" of a bat, a frog, or a bee is surely more different still from that of a human. What attributes *will* be perceived given the ability to perceive them is undoubtedly determined by many factors having to do with the functional needs of the knower interacting with the physical and social environment. One influence on how attributes will be defined by humans is clearly the category system already existent in the culture at a given time. Thus, our segmentation of a bird's body such that there is an attribute called "wings" may be influenced not only by perceptual factors such as the gestalt laws of form that would lead us to consider the wings as a separate part (Palmer 1977) but also by the fact that at present we already have a cultural and linguistic category called "birds." Viewing attributes as, at least in part, constructs of the perceiver does not negate the higher-order structural fact about attributes at issue, namely that the attributes of wings and that of feathers do co-occur in the perceived world.

These two basic principles of categorization, a drive toward cognitive economy combined with structure in the perceived world, have implications both for the level of abstraction of categories formed in a culture and for the internal structure of those categories once formed.

For purposes of explication, we may conceive of category systems as having both a vertical and horizontal dimension. The vertical dimension concerns the level of inclusiveness of the category—the dimension along which the terms collie, dog, mammal, animal, and living thing vary. The horizontal dimension concerns the segmentation of categories at the same level of inclusiveness—the dimension on which dog, cat, car, bus, chair, and sofa vary. The implication of the two principles of categorization for the vertical dimension is that not all possible levels of categorization are equally good or useful; rather, the most basic level of categorization will be the most inclusive (abstract) level at which the categories can mirror the structure of attributes perceived in the world. The implication of the principles of categorization for the horizontal dimension is that to increase the distinctiveness and flexibility of categories, categories tend to become defined in terms of prototypes or prototypical instances that contain

the attributes most representative of items inside and least representative of items outside the category.

The Vertical Dimension of Categories: Basic-Level Objects

In a programmatic series of experiments, we have attempted to argue that categories within taxonomies of concrete objects are structured such that there is generally one level of abstraction at which the most basic category cuts can be made (Rosch et al. 1976a). By *category* is meant a number of objects that are considered equivalent. Categories are generally designated by names (e.g., *dog, animal*). A *taxonomy* is a system by which categories are related to one another by means of class inclusion. The greater the inclusiveness of a category within a taxonomy, the higher the level of abstraction. Each category within a taxonomy is entirely included within one other category (unless it is the highest level category) but is not exhaustive of that more inclusive category (see Kay 1971). Thus the term *level of abstraction* within a taxonomy refers to a particular level of inclusiveness. A familiar taxonomy is the Linnean system for the classification of animals.

Our claims concerning a basic level of abstraction can be formalized in terms of cue validity (Rosch et al. 1976a) or in terms of the set theoretic representation of similarity provided by Tversky (1977, and Tversky and Gati 1978). Cue validity is a probabilistic concept; the validity of a given cue x as a predictor of a given category y (the conditional probability of y/x) increases as the frequency with which cue x is associated with category y increases and decreases as the frequency with which cue x is associated with categories other than y increases (Beach 1964a, 1964b; Reed 1972). The cue validity of an entire category may be defined as the summation of the cue validities for that category of each of the attributes of the category. A category with high cue validity is, by definition, more differentiated from other categories than one of lower cue validity. The elegant formulization that Tversky (1978) provides is in terms of the variable "category resemblance," which is defined as the weighted sum of the measures of all of the common features within a category minus the sum of the measures of all of the distinctive features. Distinctive features include those that belong to only some members of a given category as well as those belonging to contrasting categories. Thus Tversky's formalization does not weight the effect of contrast categories as much as does the cue validity formulation. Tversky suggests that two disjoint classes tend to be combined whenever the weight of the added common features exceeds the weight of the distinctive features.

A working assumption of the research on basic objects is that (1) in the perceived world, information-rich bundles of perceptual and functional attributes occur that form natural discontinuities, and that (2) basic cuts in categorization are made at these discontinuities. Suppose that basic objects (e.g., chair, car) are at the most inclusive level at which there are attributes common to all or most members of the category. Then both total cue validities and category resemblance are maximized at that level of abstraction at which basic objects are categorized. This is, categories one level more abstract will be superordinate categories (e.g., furniture, vehicle) whose members share only a few attributes

among each other. Categories below the basic level will be bundles of common and, thus, predictable attributes and functions but contain many attributes that overlap with other categories (for example, kitchen chair shares most of its attributes with other kinds of chairs).

Superordinate categories have lower total cue validity and lower category resemblance than do basic-level categories, because they have fewer common attributes; in fact, the category resemblance measure of items within the superordinate can even be negative due to the high ratio of distinctive to common features. Subordinate categories have lower total cue validity than do basic categories, because they also share most attributes with contrasting subordinate categories; in Tversky's terms, they tend to be combined because the weight of the added common features tends to exceed the weight of the distinctive features. That basic objects are categories at the level of abstraction that maximizes cue validity and maximizes category resemblance is another way of asserting that basic objects are the categories that best mirror the correlational structure of the environment.

We chose to look at concrete objects because they appeared to be a domain that was at once an indisputable aspect of complex natural language classifications yet at the same time was amenable to methods of empirical analysis. In our investigations of basic categories, the correlational structure of concrete objects was considered to consist of a number of inseparable aspects of form and function, any one of which could serve as the starting point for analysis. Four investigations provided converging operational definitions of the basic level of abstraction: attributes in common, motor movements in common, objective similarity in shape, and identifiability of averaged shapes.

Common Attributes

Ethnobiologists had suggested on the basis of linguistic criteria and field observation that the folk genus was the level of classification at which organisms had bundles of attributes in common and maximum discontinuity between classes (see Berlin 1978). The purpose of our research was to provide a systematic empirical study of the co-occurrence of attributes in the most common taxonomies of biological and man-made objects in our own culture.

The hypothesis that basic level objects are the most inclusive level of classification at which objects have numbers of attributes in common was tested for categories at three levels of abstraction for nine taxonomies: tree, bird, fish, fruit, musical instruments, tool, clothing, furniture, and vehicle. Examples of the three levels for one biological and one nonbiological taxonomy are shown in table 10.1. Criteria for choice of these specific items were that the taxonomies contain the most common (defined by word frequency) categories of concrete nouns in English, that the levels of abstraction bear simple class-inclusion relations to each other, and that those class-inclusion relations be generally known to our subjects (be agreed upon by a sample of native English speakers). The middle level of abstraction was the hypothesized basic level: For nonbiological taxonomies, this corresponded to the intuition of the experimenters (which also turned out to be consistent with Berlin's linguistic criteria); for biological categories, we assumed that the basic level would be the level of the folk generic.

Table 10.1
Examples of taxonomies used in basic object research

Superordinate	Basic level	Subordinate
Furniture	Chair	Kitchen chair
		Living-room chair
	Table	Kitchen table
		Dining-room table
	Lamp	Floor lamp
		Desk lamp
Tree	Oak	White oak
		Red oak
	Maple	Silver maple
		Sugar maple
	Birch	River birch
		White birch

Subjects received sets of words taken from these nine taxonomies; the subject's task was to list all of the attributes he could think of that were true of the items included in the class of things designated by each object name. Thus, for purposes of this study, attributes were defined operationally as whatever subjects agreed them to be with no implications for whether such analysis of an object could or could not be perceptually considered prior to knowledge of the object itself. Results of the study were as predicted: Very few attributes were listed for the superordinate categories, a significantly greater number listed for the supposed basic-level objects, and not significantly more attributes listed for subordinate-level objects than for basic-level. An additional study showed essentially the same attributes listed for visually present objects as for the object names. The single unpredicted result was that for the three biological taxonomies, the basic level, as defined by numbers of attributes in common, did not occur at the level of the folk generic but appeared at the level we had originally expected to be superordinate (e.g., *tree* rather than *oak*).

Motor Movements

Inseparable from the perceived attributes of objects are the ways in which humans habitually use or interact with those objects. For concrete objects, such interactions take the form of motor movements. For example, when performing the action of sitting down on a chair, a sequence of body and muscle movements are typically made that are inseparable from the nature of the attributes of chairs—legs, seat, back, etc. This aspect of objects is particularly important in light of the role that sensory-motor interaction with the world appears to play in the development of thought (Bruner, Olver, and Greenfield 1966; Nelson 1974; Piaget 1952).

In our study of motor movements, each of the sets of words used in the previous experiment was administered to new subjects. A subject was asked to describe, in as much finely analyzed detail as possible, the sequences of motor movements he made when using or interacting with the object. Tallies of

agreed upon listings of the same movements of the same body part in the same part of the movement sequence formed the unit of analysis. Results were identical to those of the attribute listings; basic objects were the most general classes to have motor sequences in common. For example, there are few motor programs we carry out to items of furniture in general and several specific motor programs carried out in regard to sitting down on chairs, but we sit on kitchen and living-room chairs using essentially the same motor programs.

Similarity in Shapes

Another aspect of the meaning of a class of objects is the appearance of the objects in the class. In order to be able to analyze correlational structures by different but converging methods, it was necessary to find a method of analyzing similarity in the visual aspects of the objects that was not dependent on subjects' descriptions, that was free from effects of the object's name (which would not have been the case for subjects' ratings of similarity), and that went beyond similarity of analyzable, listable attributes that had already been used in the first study described. For this purpose, outlines of the shape of two-dimensional representations of objects were used, an integral aspect of natural forms. Similarity in shape was measured by the amount of overlap of the two outlines when the outlines (normalized for size and orientation) were juxtaposed.

Results showed that the ratio of overlapped to nonoverlapped area when two objects from the same basic-level category (e.g., two cars) were superimposed was far greater than when two objects from the same superordinate category were superimposed (e.g., a car and a motorcycle). Although some gain in ratio of overlap to nonoverlap also occurred for subordinate category objects (e.g., two sports cars), the gain obtained by shifting from basic-level to subordinate objects was significantly less than the gain obtained by shifting from superordinate to basic-level objects.

Identifiability of Averaged Shapes

If the basic level is the most inclusive level at which shapes of objects of a class are similar, a possible result of such similarity may be that the basic level is also the most inclusive level at which an averaged shape of an object can be recognized. To test this hypothesis, the same normalized superimposed shapes used in the previous experiment were used to draw an averaged outline of the overlapped figures. Subjects were then asked to identify both the superordinate category and the specific object depicted. Results showed that basic objects were the most general and inclusive categories at which the objects depicted could be identified. Furthermore, overlaps of subordinate objects were no more identifiable than objects at the basic level.

In summary, our four converging operational definitions of basic objects all indicated the same level of abstraction to be basic in our taxonomies. Admittedly, the basic level for biological objects was not that predicted by the folk genus; however, this fact appeared to be simply accounted for by our subjects' lack of knowledge of the additional depth of real-world attribute structure available at the level of the folk generic (see Rosch et al. 1976a).

Implications for Other Fields

The foregoing theory of categorization and basic objects has implications for several traditional areas of study in psychology; some of these have been tested.

Imagery

The fact that basic-level objects were the most inclusive categories at which an averaged member of the category could be identified suggested that basic objects might be the most inclusive categories for which it was possible to form a mental image isomorphic to the appearance of members of the class as a whole. Experiments using a signal-detection paradigm and a priming paradigm, both of which have been previously argued to be measures of imagery (Peterson and Graham 1974; Rosch 1975c), verified that, in so far as it was meaningful to use the term *imagery*, basic objects appeared to be the most abstract categories for which an image could be reasonably representative of the class as a whole.

Perception

From all that has been said of the nature of basic classifications, it would hardly be reasonable to suppose that in perception of the world, objects were first categorized either at the most abstract or at the most concrete level possible. Two separate studies of picture verification (Rosch et al. 1976a; Smith, Balzano, and Walker 1978) indicate that, in fact, objects may be first seen or recognized as members of their basic category, and that only with the aid of additional processing can they be identified as members of their superordinate or subordinate category.

Development

We have argued that classification into categories at the basic level is overdetermined because perception, motor movements, functions, and iconic images would all lead to the same level of categorization. Thus basic objects should be the first categorizations of concrete objects made by children. In fact, for our nine taxonomies, the basic level was the first named. And even when naming was controlled, pictures of several basic-level objects were sorted into groups "because they were the same type of thing" long before such a technique of sorting has become general in children.

Language

From all that has been said, we would expect the most useful and, thus, most used name for an item to be the basic-level name. In fact, we found that adults almost invariably named pictures of the subordinate items of the nine taxonomies at the basic level, although they knew the correct superordinate and subordinate names for the objects. On a more speculative level, in the evolution of languages, one would expect names to evolve first for basic-level objects, spreading both upward and downward as taxonomies increased in depth. Of great relevance for this hypothesis are Berlin's (1972) claims for such a pattern for the evolution of plant names, and our own (Rosch et al. 1976a) and Newport and Bellugi's (1978) finding for American Sign Language of the Deaf, that

it was the basic-level categories that were most often coded by single signs and super- and subordinate categories that were likely to be missing. Thus a wide range of converging operations verify as basic the same levels of abstraction.

The Horizontal Dimension: Internal Structure of Categories: Prototypes

Most, if not all, categories do not have clear-cut boundaries. To argue that basic object categories follow clusters of perceived attributes is not to say that such attribute clusters are necessarily discontinuous.

In terms of the principles of categorization proposed earlier, cognitive economy dictates that categories tend to be viewed as being as separate from each other and as clear-cut as possible. One way to achieve this is by means of formal, necessary and sufficient criteria for category membership. The attempt to impose such criteria on categories marks virtually all definitions in the tradition of Western reason. The psychological treatment of categories in the standard concept-identification paradigm lies within this tradition. Another way to achieve separateness and clarity of actually continuous categories is by conceiving of each category in terms of its clear cases rather than its boundaries. As Wittgenstein (1953) has pointed out, categorical judgments become a problem only if one is concerned with boundaries—in the normal course of life, two neighbors know on whose property they are standing without exact demarcation of the boundary line. Categories can be viewed in terms of their clear cases if the perceiver places emphasis on the correlational structure of perceived attributes such that the categories are represented by their most structured portions.

By prototypes of categories we have generally meant the clearest cases of category membership defined operationally by people's judgments of goodness of membership in the category. A great deal of confusion in the discussion of prototypes has arisen from two sources. First, the notion of prototypes has tended to become reified as though it meant a specific category member or mental structure. Questions are then asked in an either-or fashion about whether something is or is not the prototype or part of the prototype in exactly the same way in which the question would previously have been asked about the category boundary. Such thinking precisely violates the Wittgensteinian insight that we can judge how clear a case something is and deal with categories on the basis of clear cases in the total absence of information about boundaries. Second, the empirical findings about prototypicality have been confused with theories of processing—that is, there has been a failure to distinguish the structure of categories from theories concerning the use of that structure in processing. Therefore, let us first attempt to look at prototypes in as purely structural a fashion as possible. We will focus on what may be said about prototypes based on operational definitions and empirical findings alone without the addition of processing assumptions.

Perception of typicality differences is, in the first place, an empirical fact of people's judgments about category membership. It is by now a well-documented finding that subjects overwhelmingly agree in their judgments of how good an example or clear a case members are of a category, even for categories about whose boundaries they disagree (Rosch 1974, 1975b). Such

judgments are reliable even under changes of instructions and items (Rips, Shoben, and Smith 1973; Rosch 1975b, 1975c; Rosch and Mervis 1975). Were such agreement and reliability in judgment not to have been obtained, there would be no further point in discussion or investigation of the issue. However, given the empirical verification of degree of prototypicality, we can proceed to ask what principles determine which items will be judged the more prototypical and what other variables might be affected by prototypicality.

In terms of the basic principles of category formation, the formation of category prototypes should, like basic levels of abstraction, be determinate and be closely related to the initial formation of categories. For categories of concrete objects (which do not have a physiological basis, as categories such as colors and forms apparently do—Rosch 1974), a reasonable hypothesis is that prototypes develop through the same principles such as maximization of cue validity and maximization of category resemblance[1] as those principles governing the formation of the categories themselves.

In support of such a hypothesis, Rosch and Mervis (1975) have shown that the more prototypical of a category a member is rated, the more attributes it has in common with other members of the category and the fewer attributes in common with members of the contrasting categories. This finding was demonstrated for natural language superordinate categories, for natural language basic-level categories, and for artificial categories in which the definition of attributes and the amount of experience with items was completely specified and controlled. The same basic principles can be represented in ways other than through attributes in common. Because the present theory is a structural theory, one aspect of it is that centrality shares the mathematical notions inherent in measures like the mean and mode. Prototypical category members have been found to represent the means of attributes that have a metric, such as size (Reed 1972; Rosch, Simpson, and Miller 1976).

In short, prototypes appear to be just those members of a category that most reflect the redundancy structure of the category as a whole. That is, if categories form to maximize the information-rich cluster of attributes in the environment and, thus, the cue validity or category resemblance of the attributes of categories, prototypes of categories appear to form in such a manner as to maximize such clusters and such cue validity still further within categories.

It is important to note that for natural language categories both at the superordinate and basic levels, the extent to which items have attributes common to the category was highly negatively correlated with the extent to which they have attributes belonging to members of contrast categories. This appears to be part of the structure of real-world categories. It may be that such structure is given by the correlated clusters of attributes of the real world. Or such structure, may be a result of the human tendency once a contrast exists to define attributes for contrasting categories so that the categories will be maximally distinctive. In either case, it is a fact that both representativeness within a category and distinctiveness from contrast categories are correlated with prototypicality in real categories. For artificial categories, either principle alone will produce prototype effects (Rosch et al. 1976b; Smith and Balzano, personal communication) depending on the structure of the stimulus set. Thus to perform experiments to try to distinguish which principle is the *one* that deter-

mines prototype formation and category processing appears to be an artificial exercise.

Effects of Prototypicality on Psychological Dependent Variables

The fact that prototypicality is reliably rated and is correlated with category structure does not have clear implications for particular processing models nor for a theory of cognitive representations of categories (see the introduction to Part III of Rosch and Lloyd 1978 and Palmer 1978). What is very clear from the extant research is that the prototypicality of items within a category can be shown to affect virtually all of the major dependent variables used as measures in psychological research.

Speed of Processing: Reaction Time

The speed with which subjects can judge statements about category membership is one of the most widely used measures of processing in semantic memory research within the human information-processing framework. Subjects typically are required to respond true or false to statements of the form: *X* item is a member of *Y* category, where the dependent variable of interest is reaction time. In such tasks, for natural language categories, responses of true are invariably faster for the items that have been rated more prototypical. Furthermore, Rosch et al. (1976b) had subjects learn artificial categories where prototypicality was defined structurally for some subjects in terms of distance of a gestalt configuration from a prototype, for others in terms of means of attributes, and for still others in terms of family resemblance between attributes. Factors other than the structure of the category, such as frequency, were controlled. After learning was completed, reaction time in a category membership verification task proved to be a function of structural prototypicality.

Speed of Learning of Artificial Categories (Errors) and Order of Development in Children

Rate of learning of new material and the naturally obtainable measure of learning (combined with maturation) reflected in developmental order are two of the most pervasive dependent variables in psychological research. In the artificial categories used by Rosch et al. (1976b), prototypicality for all three types of stimulus material predicted speed of learning of the categories. Developmentally, Anglin (1976) obtained evidence that young children learn category membership of good examples of categories before that of poor examples. Using a category-membership verification technique, Rosch (1973) found that the differences in reaction time to verify good and poor members were far more extreme for 10-year-old children than for adults, indicating that the children had learned the category membership of the prototypical members earlier than that of other members.

Order and Probability of Item Output

Item output is normally taken to reflect some aspect of storage, retrieval, or category search. Battig and Montague (1969) provided a normative study of the probability with which college students listed instances of superordinate

semantic categories. The order is correlated with prototypicality ratings (Rosch 1975b). Furthermore, using the artificial categories in which frequency of experience with all items was controlled, Rosch et al. (1976b) demonstrated that the most prototypical items were the first and most frequently produced items when subjects were asked to list the members of the category.

Effects of Advance Information on Performance: Set, Priming
For colors (Rosch 1975c), for natural superordinate semantic categories (Rosch 1975b), and for artificial categories (Rosch et al. 1976b), it has been shown that degree of prototypicality determines whether advance information about the category name facilitates or inhibits responses in a matching task.

The Logic of Natural Language Use of Category Terms: Hedges, Substitutability into Sentences, Superordination in ASL
Although logic may treat categories as though membership is all or none, natural languages themselves possess linguistic mechanisms for coding and coping with gradients of category membership.

1. *Hedges*. In English there are qualifying terms such as "almost" and "virtually," which Lakoff (1972) calls "hedges." Even those who insist that statements such as "A robin is a bird" and "A penguin is a bird" are equally true, have to admit different hedges applicable to statements of category membership. Thus it is correct to say that a penguin is technically a bird but not that a robin is technically a bird, because a robin is more than just technically a bird; it is a real bird, a bird par excellence. Rosch (1975a) showed that when subjects were given sentence frames such as "*X* is virtually *Y*," they reliably placed the more prototypical member of a pair of items into the referent slot, a finding which is isomorphic to Tversky's work on asymmetry of similarity relations (Tversky & Gati 1978).
2. *Substitutability into sentences*. The meaning of words is intimately tied to their use in sentences. Rosch (1977) has shown that prototypicality ratings for members of superordinate categories predict the extent to which the member term is substitutable for the superordinate word in sentences. Thus, in the sentence "Twenty or so birds often perch on the telephone wires outside my window and twitter in the morning," the term "sparrow" may readily be substituted for "bird" but the result turns ludicrous by substitution of "turkey," an effect which is not simply a matter of frequency (Rosch 1975d).
3. *Productive superordinates in ASL*. Newport and Bellugi (1978) demonstrate that when superordinates in ASL are generated by means of a partial fixed list of category members, those members are the more prototypical items in the category.

In summary, evidence has been presented that prototypes of categories are related to the major dependent variables with which psychological processes are typically measured. What the work summarized does not tell us, however, is considerably more than it tells us. The pervasiveness of prototypes in real-

world categories and of prototypicality as a variable indicates that prototypes must have some place in psychological theories of representation, processing, and learning. However, prototypes themselves do not constitute any particular model of processes, representations, or learning. This point is so often misunderstood that it requires discussion:

1. To speak of *a prototype* at all is simply a convenient grammatical fiction; what is really referred to are judgments of degree of prototypicality. Only in some artificial categories is there by definition a literal single prototype (for example, Posner, Goldsmith, and Welton 1967; Reed 1972; Rosch et al. 1976b). For natural-language categories, to speak of a single entity that is the prototype is either a gross misunderstanding of the empirical data or a covert theory of mental representation.
2. Prototypes do not constitute any particular processing model for categories. For example, in pattern recognition, as Palmer (1978) points out, a prototype can be described as well by feature lists or structural descriptions as by templates. And many different types of matching operations can be conceived for matching to a prototype given any of these three modes of representation of the prototypes. Other cognitive processes performed on categories such as verifying the membership of an instance in a category, searching the exemplars of a category for the member with a particular attribute, or understanding the meaning of a paragraph containing the category name are not bound to any single process model by the fact that we may acknowledge prototypes. What the facts about prototypicality do contribute to processing notions is a constraint—process models should not be inconsistent with the known facts about prototypes. For example, a model should not be such as to predict equal verification times for good and bad examples of categories nor predict completely random search through a category.
3. Prototypes do not constitute a theory of representation of categories. Although we have suggested elsewhere that it would be reasonable in light of the basic principles of categorization, if categories were represented by prototypes that were most representative of the items in the category and least representative of items outside the category (Rosch and Mervis 1975; Rosch 1977), such a statement remains an unspecified formula until it is made concrete by inclusion in some specific theory of representation. For example, different theories of semantic memory can contain the notion of prototypes in different fashions (Smith, 1978). Prototypes can be represented either by propositional or image systems (see Kosslyn 1978 and Palmer 1978). As with processing models, the facts about prototypes can only constrain, but do not determine, models of representation. A representation of categories in terms of conjoined necessary and sufficient attributes alone would probably be incapable of handling all of the presently known facts, but there are many representations other than necessary and sufficient attributes that are possible.
4. Although prototypes must be learned, they do not constitute any particular theory of category learning. For example, learning of prototypicality

in the types of categories examined in Rosch and Mervis (1975) could be represented in terms of counting attribute frequency (as in Neuman 1974), in terms of storage of a set of exemplars to which one later matched the input (see Shepp 1978 and the introduction to Part II of Rosch and Lloyd 1978), or in terms of explicit teaching of the prototypes once prototypicality within a category is established in a culture (e.g., "Now that's a *real* coat.")

In short, prototypes only constrain but do not specify representation and process models. In addition, such models further constrain each other. For example, one could not argue for a frequency count of attributes in children's learning of prototypes of categories if one had reason to believe that children's representation of attributes did not allow for separability and selective attention to each attribute (see Garner 1978 and the introduction to Part II of Rosch and Lloyd 1978).

Two Problematical Issues

The Nature of Perceived Attributes

The derivations of basic objects and of prototypes from the basic principles of categorization have depended on the notion of a structure in the perceived world—bundles of perceived world attributes that formed natural discontinuities. When the research on basic objects and their prototypes was initially conceived (Rosch et al. 1976a), I thought of such attributes as inherent in the real world. Thus, given an organism that had sensory equipment capable of perceiving attributes such as wings and feathers, it was a fact in the real world that wings and feathers co-occurred. The state of knowledge of a person might be ignorant of (or indifferent or inattentive to) the attributes or might know of the attributes but be ignorant concerning their correlation. Conversely, a person might know of the attributes and their correlational structure but exaggerate that structure, turning partial into complete correlations (as when attributes true only of many members of a category are thought of as true of all members). However, the environment was thought to constrain categorizations in that human knowledge could not provide correlational structure where there was none at all. For purposes of the basic object experiments, perceived attributes were operationally defined as those attributes listed by our subjects. Shape was defined as measured by our computer programs. We thus seemed to have our system grounded comfortably in the real world.

On contemplation of the nature of many of the attributes listed by our subjects, however, it appeared that three types of attributes presented a problem for such a realistic view: (1) some attributes, such as "seat" for the object "chair," appeared to have names that showed them not to be meaningful prior to knowledge of the object as chair; (2) some attributes such as "large" for the object "piano" seemed to have meaning only in relation to categorization of the object in terms of a superordinate category—piano is large for furniture but small for other kinds of objects such as buildings; (3) some attributes such as "you eat on it" for the object "table" were functional attributes that seemed to require knowledge about humans, their activities, and the real world in order

to be understood (see Miller 1978). That is, it appeared that the analysis of objects into attributes was a rather sophisticated activity that our subjects (and indeed a system of cultural knowledge) might well be considered to be able to impose only *after* the development of the category system.

In fact, the same laws of cognitive economy leading to the push toward basic-level categories and prototypes might also lead to the definition of attributes of categories such that the categories once given would appear maximally distinctive from one another and such that the more prototypical items would appear even more representative of their own and less representative of contrastive categories. Actually, in the evolution of the meaning of terms in languages, probably both the constraint of real-world factors and the construction and reconstruction of attributes are continually present. Thus, given a particular category system, attributes are defined such as to make the system appear as logical and economical as possible. However, if such a system becomes markedly out of phase with real-world constraints, it will probably tend to evolve to be more in line with those constraints—with redefinition of attributes ensuing if necessary. Unfortunately, to state the matter in such a way is to provide no clear place at which we can enter the system as analytical scientists. What is the unit with which to start our analysis? Partly in order to find a more basic real-world unit for analysis than attributes, we have turned our attention to the contexts in which objects occur—that is, to the culturally defined events in which objects serve as props.

The Role of Context in Basic-Level Objects and Prototypes

It is obvious, even in the absence of controlled experimentation, that a man about to buy a chair who is standing in a furniture store surrounded by different chairs among which he must choose will think and speak about chairs at other than the basic level of "chair." Similarly, in regard to prototypes, it is obvious that if asked for the most typical African animal, people of any age will not name the same animal as when asked for the most typical American pet animal. Because interest in context is only beginning, it is not yet clear just what experimentally defined contexts will affect what dependent variables for what categories. But it is predetermined that there will be context effects for both the level of abstraction at which an object is considered and for which items are named, learned, listed, or expected in a category. Does this mean that our findings in regard to basic levels and prototypes are relevant only to the artificial situation of the laboratory in which a context is not specified?

Actually, both basic levels and prototypes are, in a sense, theories about context itself. The basic level of abstraction is that level of abstraction that is appropriate for using, thinking about, or naming an object in most situations in which the object occurs (Rosch et al. 1976a). And when a context is not specified in an experiment, people must contribute their own context. Presumably, they do not do so randomly. Indeed, it seems likely that, in the absence of a specified context, subjects assume what they consider the normal context or situation for occurrence of that object. To make such claims about categories appears to demand an analysis of the actual events in daily life in which objects occur.

The Role of Objects in Events

The attempt we have made to answer the issues of the origin of attributes and the role of context has been in terms of the use of objects in the events of daily human life. The study of events grew out of an interest in categorizations of the flow of experience. That is, our initial interest was in the question of whether any of the principles of categorization we had found useful for understanding concrete objects appeared to apply to the cutting up of the continuity of experience into the discrete bounded temporal units that we call *events*.

Previously, events have been studied primarily from two perspectives in psychology. Within ecological and social psychology, an observer records and attempts to segment the stream of another person's behavior into event sequences (for example, Barker and Wright 1955; Newtson 1976). And within the artificial intelligence tradition. Story Understanders are being constructed that can "comprehend," by means of event scripts, statements about simple, culturally predictable sequences such as going to a restaurant (Shank 1975).

The unit of the event would appear to be a particularly important unit for analysis. Events stand at the interface between an analysis of social structure and culture and an analysis of individual psychology. It may be useful to think of scripts for events as the level of theory at which we can specify how culture and social structure enter the individual mind. Could we use events as the basic unit from which to derive an understanding of objects? Could we view objects as props for the carrying out of events and have the functions, perceptual attributes, and levels of abstraction of objects fall out of their role in such events?

Our research to date has been a study rather than an experiment and more like a pilot study at that. Events were defined neither by observation of others nor by a priori units for scripts but introspectively in the following fashion. Students in a seminar on events were asked to choose a particular evening on which to list the events that they remembered of that day—e.g., to answer the question what did I do? (or what happened to me?) that day by means of a list of the names of the events. They were to begin in the morning. The students were aware of the nature of the inquiry and that the focus of interest was on the units that they would perceive as the appropriate units into which to chunk the days' happenings. After completing the list for that day, they were to do the same sort of lists for events remembered from the previous day, and thus to continue backwards to preceding days until they could remember no more day's events. They also listed events for units smaller and larger than a day: for example, the hour immediately preceding writing and the previous school quarter.

The results were somewhat encouraging concerning the tractability of such a means of study. There was considerable agreement on the kinds of units into which a day should be broken—units such as making coffee, taking a shower, and going to statistics class. No one used much smaller units: That is, units such as picking up the toothpaste tube, squeezing toothpaste onto the brush, etc., never occurred. Nor did people use larger units such as "got myself out of the house in the morning" or "went to all my afternoon classes." Furthermore,

the units that were listed did not change in size or type with their recency or remoteness in time to the writing. Thus, for the time unit of the hour preceding writing, components of events were not listed. Nor were larger units of time given for a day a week past than for the day on which the list was composed. Indeed, it was dramatic how, as days further and further in the past appeared, fewer and fewer events were remembered although the type of unit for those that were remembered remained the same. That is, for a day a week past, a student would not say that he now only remembered getting himself out of the house in the morning (though such "summarizing" events could be inferred); rather he either did or did not remember feeding the cat that day (an occurrence that could also be inferred but for which inference and memory were introspectively clearly distinguishable). Indeed, it appeared that events such as "all the morning chores" as a whole do not have a memory representation separate from memory of doing the individual chores—perhaps in the way that superordinate categories, such as furniture, do not appear to be imageable per se apart from imaging individual items in the category. It should be noted that event boundaries appeared to be marked in a reasonable way by factors such as changes of the actors participating with ego, changes in the objects ego interacts with, changes in place, and changes in the type or rate of activity with an object, and by notable gaps in time between two reported events.

A good candidate for the basic level of abstraction for events is the type of unit into which the students broke their days. The events they listed were just those kinds of events for which Shank (1975) has provided scripts. Scripts of events analyze the event into individual units of action; these typically occur in a predictable order. For example, the script for going to a restaurant contains script elements such as entering, going to a table, ordering, eating, and paying. Some recent research has provided evidence for the psychological reality of scripts and their elements (Bower 1976).

Our present concern is with the role of concrete objects in events. What categories of objects are required to serve as props for events at the level of abstraction of those listed by the students? In general, we found that the event name itself combined most readily with superordinate noun categories; thus, one gets dressed with clothes and needs various kitchen utensils to make breakfast. When such activities were analyzed into their script elements, the basic level appeared as the level of abstraction of objects necessary to script the events; e.g., in getting dressed, one puts on pants, sweater, and shoes, and in making breakfast, one cooks eggs in a frying pan.

With respect to prototypes, it appears to be those category members judged the more prototypical that have attributes that enable them to fit into the typical and agreed upon script elements. We are presently collecting normative data on the intersection of common events, the objects associated with those events and the other sets of events associated with those objects.[2] In addition, object names for eliciting events are varied in level of abstraction and in known prototypicality in given categories. Initial results show a similar pattern to that obtained in the earlier research in which it was found that the more typical members of superordinate categories could replace the superordinate in sentence frames generated by subjects told to "make up a sentence" that used the

superordinate (Rosch 1977). That is, the task of using a given concrete noun in a sentence appears to be an indirect method of eliciting a statement about the events in which objects play a part; that indirect method showed clearly that prototypical category members are those that can play the role in events expected of members of that category.

The use of deviant forms of object names in narratives accounts for several recently explored effects in the psychological literature. Substituting object names at other than the basic level within scripts results in obviously deviant descriptions. Substitution of superordinates produces just those types of narrative that Bransford and Johnson (1973) have claimed are not comprehended; for example, "The procedure is actually quite simple. First you arrange things into different groups. Of course, one pile may be sufficient [p. 400]." It should be noted in the present context that what Bransford and Johnson call context cues are actually names of basic-level events (e.g., washing clothes) and that one function of hearing the event name is to enable the reader to translate the superordinate terms into basic-level objects and actions. Such a translation appears to be a necessary aspect of our ability to match linguistic descriptions to world knowledge in a way that produces the "click of comprehension."

On the other hand, substitution of subordinate terms for basic-level object names in scripts gives the effect of satire or snobbery. For example, a review (Garis 1975) of a pretentious novel accused of actually being about nothing more than brand-name snobbery concludes, "And so, after putting away my 10-year-old Royal 470 manual and lining up my Mongol number 3 pencils on my Goldsmith Brothers Formica imitation-wood desk, I slide into my oversize squirrel-skin L. L. Bean slippers and shuffle off to the kitchen. There, holding *Decades* in my trembling right hand, I drop it, *plunk*, into my new Sears 20-gallon, celadon-green Permanex trash can [p. 48]."

Analysis of events is still in its initial stages. It is hoped that further understanding of the functions and attributes of objects can be derived from such an analysis.

Summary

The first part of this chapter showed how the same principles of categorization could account for the taxonomic structure of a category system organized around a basic level and also for the formation of the categories that occur within this basic level. Thus the principles described accounted for both the vertical and horizontal structure of category systems. Four converging operations were employed to establish the claim that the basic level provides the cornerstone of a taxonomy. The section on prototypes distinguished the empirical evidence for prototypes as structural facts about categories from the possible role of prototypes in cognitive processing, representation, and learning. Then we considered assumptions about the nature of the attributes of real-world objects and assumptions about context—insofar as attributes and contexts underlie the claim that there is structure in the world. Finally, a highly tentative pilot study of attributes and functions of objects as props in culturally defined events was presented.

Notes

1. Tversky formalizes prototypicality as the member or members of the category with the highest summed similarity to all members of the category. This measure, although formally more tractable than that of cue validity, does not take account, as cue validity does, of an item's dissimilarity to contrast categories. This issue is discussed further later.
2. This work is being done by Elizabeth Kreusi.

References

Anglin, J. Les premiers termes de référence de l'enfant. In S. Ehrlich and E. Tulving (Eds.), *La memoire sémantique*. Paris: Bulletin de Psychologie, 1976.

Barker, R., and Wright, H. *Midwest and its children*. Evanston, Ill.: Row-Peterson, 1955.

Battig, W. F., and Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, *80* (3, Pt. 2).

Beach, L. R. Cue probabilism and inference behavior. *Psychological Monographs*. 1964, *78* (Whole No. 582). (a)

Beach, L. R. Recognition, assimilation, and identification of objects. *Psychological Monographs*, 1964, *78* (Whole No. 583). (b)

Berlin, B. Speculations on the growth of ethnobotanical nomenclature. *Language in Society*, 1972, *1*, 51–86.

Berlin, B. Ethnobiological classification. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1978.

Borges, J. L. *Other inquisitions 1937–1952*. New York: Washington Square Press, 1966.

Bower, G. *Comprehending and recalling stories*. Paper presented as Division 3 presidential address to the American Psychological Association, Washington, D.C., September 1976.

Bransford, J. D., and Johnson, M. K. Considerations of some problems of comprehension. In W. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

Bruner, J. S., Olver, R. R., and Greenfield, P. M. *Studies in cognitive growth*. New York: Wiley, 1966.

Garis, L. The Margaret Mead of Madison Avenue. *Ms*., March 1975, pp. 47–48.

Garner, W. R. *The processing of information and structure*. New York: Wiley, 1974.

Garner, W. R. Aspects of a stimulus: Features, dimensions, and configurations. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1978.

Kay, P. Taxonomy and semantic contrast. *Language*, 1971, *47*, 866–887.

Kosslyn, S. M. Imagery and internal representation. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1978.

Lakoff, G. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Papers from the eighth regional meeting, Chicago Linguistics Society*. Chicago: University of Chicago Linguistics Department, 1972.

Miller, G. A. Practical and lexical knowledge. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1978.

Nelson, K. Concept, word and sentence: Interrelations in acquisition and development. *Psychological Review*, 1974, *81*, 267–285.

Neuman, P. G. An attribute frequency model for the abstraction of prototypes. *Memory and Cognition*, 1974, *2*, 241–248.

Newport, E. L., and Bellugi, U. Linguistic expression of category levels in a visual-gestural language: A flower is a flower is a flower. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1978.

Newtson, D. Foundations of attribution: The perception of ongoing behavior. In J. Harvey, W. Ickes, and R. Kidd (Eds.), *New directions in attribution research*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Palmer, S. Hierarchical structure in perceptual representation. *Cognitive Psychology*, 1977, *9*, 441–474.

Palmer, S. E. Fundamental aspects of cognitive representation. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1978.

Peterson, M. J., and Graham, S. E. Visual detection and visual imagery. *Journal of Experimental Psychology*, 1974, *103*, 509–514.

Piaget, J. *The origins of intelligence in children*. New York: International Universities Press, 1952.

Posner, M. I., Goldsmith, R., and Welton, K. E. Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology*, 1967, *73*, 28–38.

Reed, S. K. Pattern recognition and categorization. *Cognitive Psychology*, 1972, *3*, 382–407.

Rips, L. J., Shoben, E. J., and Smith, E. E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 1–20.

Rosch, E. On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press, 1973.

Rosch, E. Linguistic relativity. In A. Silverstein (Ed.), *Human communication: Theoretical perspectives*. New York: Halsted Press, 1974.

Rosch, E. Cognitive reference points. *Cognitive Psychology*, 1975, *7*, 532–547. (a)

Rosch, E. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 1975, *104*, 192–233. (b)

Rosch, E. The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, *1*, 303–322. (c)

Rosch, E. Universals and cultural specifics in human categorization. In R. Brislin, S. Bochner, and W. Lonner (Eds.), *Cross-cultural perspectives on learning*. New York: Halsted Press, 1975. (d)

Rosch, E. Human categorization. In N. Warren (Ed.), *Advances in cross-cultural psychology* (Vol. 1). London: Academic Press, 1977.

Rosch, E., and Lloyd, B. B. *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1978.

Rosch, E., and Mervis, C. B. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 1975, *7*, 573–605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyes-Braem, P. Basic objects in natural categories. *Cognitive Psychology*, 1976, *8*, 382–439. (a)

Rosch, E., Simpson, C., and Miller, R. S. Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*. 1976, *2*, 491–502. (b)

Shank, R. C. The structure of episodes in memory. In D. G. Bobrow and A. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

Shepp, B. E. From perceived similarity to dimensional structure: A new hypothesis about perspective development. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1978.

Smith, E. E. Theories of semantic memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Smith, E. E., and Balzano, G. J. Personal communication, April 1977.

Smith, E. E., Balzano, G. J., and Walker, J. H. Nominal, perceptual, and semantic codes in picture categorization. In J. Cotton and R. Klatzky (Eds.), *Semantic factors in cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Tversky, A. Features of similarity. *Psychological Review*, 1977, *84*, 327–352.

Tversky, A., and Gati, I. Studies of similarity. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1978.

Wittgenstein, L. *Philosophical investigations*. New York: Macmillan, 1953.

Chapter 11

Philosophical Investigations, Sections 65–78

Ludwig Wittgenstein

---

65. Here we come up against the great question that lies behind all these considerations.—For someone might object against me: "You take the easy way out! You talk about all sorts of language-games, but have nowhere said what the essence of a language-game, and hence of language, is: what is common to all these activities, and what makes them into language or parts of language. So you let yourself off the very part of the investigation that once gave you yourself most headache, the part about the *general form of propositions* and of language."

And this is true.—Instead of producing something common to all that we call language, I am saying that these phenomena have no one thing in common which makes us use the same word for all,—but that they are *related* to one another in many different ways. And it is because of this relationship, or these relationships, that we call them all "language." I will try to explain this.

66. Consider for example the proceedings that we call "games." I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all?—Don't say: "There *must* be something common, or they would not be called 'games'"—but *look and see* whether there is anything common to all.—For if you look at them you will not see something that is common to *all*, but similarities, relationships, and a whole series of them at that. To repeat: don't think, but look!—Look for example at board-games, with their multifarious relationships. Now pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball-games, much that is common is retained, but much is lost.—Are they all 'amusing'? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience. In ball-games there is winning and losing; but when a child throws his ball at the wall and catches it again, this feature has disappeared. Look at the parts played by skill and luck; and at the difference between skill in chess and skill in tennis. Think now of games like ring-a-ring-a-roses; here is the element of amusement, but how many other characteristic features have disappeared! And we can go through the many, many other groups of games in the same way; can see how similarities crop up and disappear.

And the result of this examination is: we see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail.

From chapter 6 in *Concepts: Core Readings*, ed. E. Margolis and S. Laurence (Cambridge, MA: MIT Press, 1999), 171–174. Reprinted with permission.

67. I can think of no better expression to characterize these similarities than "family resemblances"; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. overlap and criss-cross in the same way.—And I shall say: 'games' form a family.

And for instance the kinds of number form a family in the same way. Why do we call something a "number"? Well, perhaps because it has a—direct—relationship with several things that have hitherto been called number; and this can be said to give it an indirect relationship to other things we call the same name. And we extend our concept of number as in spinning a thread we twist fibre on fibre. And the strength of the thread does not reside in the fact that some one fibre runs through its whole length, but in the overlapping of many fibres.

But if someone wished to say: "There is something common to all these constructions—namely the disjunction of all their common properties"—I should reply: Now you are only playing with words. One might as well say: "Something runs through the whole thread—namely the continuous overlapping of those fibres."

68. "All right: the concept of number is defined for you as the logical sum of these individual interrelated concepts: cardinal numbers, rational numbers, real numbers, etc.; and in the same way the concept of a game as the logical sum of a corresponding set of sub-concepts."—It need not be so. For I *can* give the concept 'number' rigid limits in this way, that is, use the word "number" for a rigidly limited concept, but I can also use it so that the extension of the concept is *not* closed by a frontier. And this is how we do use the word "game." For how is the concept of a game bounded? What still counts as a game and what no longer does? Can you give the boundary? No. You can *draw* one; for none has so far been drawn. (But that never troubled you before when you used the word "game.")

"But then the use of the word is unregulated, the 'game' we play with it is unregulated."—It is not everywhere circumscribed by rules; but no more are there any rules for how high one throws the ball in tennis, or how hard; yet tennis is a game for all that and has rules too.

69. How should we explain to someone what a game is? I imagine that we should describe *games* to him, and we might add: "This *and similar things* are called 'games.'" And do we know any more about it ourselves? Is it only other people whom we cannot tell exactly what a game is?—But this is not ignorance. We do not know the boundaries because none have been drawn. To repeat, we can draw a boundary—for a special purpose. Does it take that to make the concept usable? Not at all! (Except for that special purpose.) No more than it took the definition: 1 pace = 75 cm. to make the measure of length 'one pace' usable. And if you want to say "But still, before that it wasn't an exact measure," then I reply: very well, it was an inexact one.—Though you still owe me a definition of exactness.

70. "But if the concept 'game' is uncircumscribed like that, you don't really know what you mean by a 'game.'"—When I give the description: "The ground was quite covered with plants"—do you want to say I don't know what I am talking about until I can give a definition of a plant?

My meaning would be explained by, say, a drawing and the words "The ground looked roughly like this." Perhaps I even say "it looked *exactly* like this."—Then were just *this* grass and *these* leaves there, arranged just like this? No, that is not what it means. And I should not accept any picture as exact in *this* sense.

Someone says to me: "Shew the children a game." I teach them gaming with dice, and the other says "I didn't mean that sort of game." Must the exclusion of the game with dice have come before his mind when he gave me the order?

71. One might say that the concept 'game' is a concept with blurred edges.—"But is a blurred concept a concept at all?"—Is an indistinct photograph a picture of a person at all? Is it even always an advantage to replace an indistinct picture by a sharp one? Isn't the indistinct one often exactly what we need?

Frege compares a concept to an area and says that an area with vague boundaries cannot be called an area at all. This presumably means that we cannot do anything with it.—But is it senseless to say: "Stand roughly there"? Suppose that I were standing with someone in a city square and said that. As I say it I do not draw any kind of boundary, but perhaps point with my hand—as if I were indicating a particular *spot*. And this is just how one might explain to someone what a game is. One gives examples and intends them to be taken in a particular way.—I do not, however, mean by this that he is supposed to see in those examples that common thing which I—for some reason—was unable to express; but that he is now to *employ* those examples in a particular way. Here giving examples is not an *indirect* means of explaining—in default of a better. For any general definition can be misunderstood too. The point is that *this* is how we play the game. (I mean the language-game with the word "game.")

72. *Seeing what is common*. Suppose I shew someone various multicoloured pictures, and say: "The colour you see in all these is called 'yellow ochre.'"—This is a definition, and the other will get to understand it by looking for and seeing what is common to the pictures. Then he can look *at*, can point *to*, the common thing.

Compare with this a case in which I shew him figures of different shapes all painted the same colour, and say: "What these have in common is called 'yellow ochre.'"

And compare this case: I shew him samples of different shades of blue and say: "The colour that is common to all these is what I call 'blue.'"

73. When someone defines the names of colours for me by pointing to samples and saying "This colour is called 'blue,' this 'green' . . ." this case can be compared in many respects to putting a table in my hands, with the words written under the colour-samples.—Though this comparison may mislead in many ways.—One is now inclined to extend the comparison: to have understood the definition means to have in one's mind an idea of the thing defined, and that is a sample or picture. So if I am shewn various different leaves and told "This is called a 'leaf,'" I get an idea of the shape of a leaf, a picture of it in my mind.—But what does the picture of a leaf look like when it does not shew us any particular shape, but 'what is common to all shapes of leaf'? Which shade is the 'sample in my mind' of the colour green—the sample of what is common to all shades of green?

"But might there not be such 'general' samples? Say a schematic leaf, or a sample of *pure* green?"—Certainly there might. But for such a schema to be understood as a *schema*, and not as the shape of a particular leaf, and for a slip of pure green to be understood as a sample of all that is greenish and not as a sample of pure green—this in turn resides in the way the samples are used.

Ask yourself: what *shape* must the sample of the colour green be? Should it be rectangular? Or would it then be the sample of a green rectangle?—So should it be 'irregular' in shape? And what is to prevent us then from regarding it—that is, from using it—only as a sample of irregularity of shape?

74. Here also belongs the idea that if you see this leaf as a sample of 'leaf shape in general' you *see* it differently from someone who regards it as, say, a sample of this particular shape. Now this might well be so—though it is not so—for it would only be to say that, as a matter of experience, if you *see* the leaf in a particular way, you use it in such-and-such a way or according to such-and-such rules. Of course, there is such a thing as seeing in *this* way or *that*; and there are also cases where whoever sees a sample like *this* will in general use it in *this* way, and whoever sees it otherwise in another way. For example, if you see the schematic drawing of a cube as a plane figure consisting of a square and two rhombi you will, perhaps, carry out the order "Bring me something like this" differently from someone who sees the picture three-dimensionally.

75. What does it mean to know what a game is? What does it mean, to know it and not be able to say it? Is this knowledge somehow equivalent to an unformulated definition? So that if it were formulated I should be able to recognize it as the expression of my knowledge? Isn't my knowledge, my concept of a game, completely expressed in the explanations that I could give? That is, in my describing examples of various kinds of games; shewing how all sorts of other games can be constructed on the analogy of these; saying that I should scarcely include this or this among games; and so on.

76. If someone were to draw a sharp boundary I could not acknowledge it as the one that I too always wanted to draw, or had drawn in my mind. For I did not want to draw one at all. His concept can then be said to be not the same as mine, but akin to it. The kinship is that of two pictures, one of which consists of colour patches with vague contours, and the other of patches similarly shaped and distributed, but with clear contours. The kinship is just as undeniable as the difference.

77. And if we carry this comparison still further it is clear that the degree to which the sharp picture *can* resemble the blurred one depends on the latter's degree of vagueness. For imagine having to sketch a sharply defined picture 'corresponding' to a blurred one. In the latter there is a blurred red rectangle: for it you put down a sharply defined one. Of course—several such sharply defined rectangles can be drawn to correspond to the indefinite one.—But if the colours in the original merge without a hint of any outline won't it become a hopeless task to draw a sharp picture corresponding to the blurred one? Won't you then have to say: "Here I might just as well draw a circle or heart as a rectangle, for all the colours merge. Anything—and nothing—is right."—— And this is the position you are in if you look for definitions corresponding to our concepts in aesthetics or ethics.

In such a difficulty always ask yourself: How did we *learn* the meaning of this word ("good" for instance)? From what sort of examples? in what language-games? Then it will be easier for you to see that the word must have a family of meanings.

78. Compare *knowing* and *saying*:

how many feet high Mont Blanc is—
how the word "game" is used—
how a clarinet sounds.

If you are surprised that one can know something and not be able to say it, you are perhaps thinking of a case like the first. Certainly not of one like the third.

Chapter 12

The Exemplar View

Edward E. Smith and Douglas L. Medin

In this chapter we take up our third view of concepts, the exemplar view. Since this view is quite new and has not been extensively developed, we will not give separate treatments of featural, dimensional, and holistic approaches. Instead, we will sometimes rely on featural descriptions, other times on dimensional ones.

Rationale for the Exemplar View

As its name suggests, the exemplar view holds that concepts are represented by their exemplars (at least in part) rather than by an abstract summary. This idea conflicts not only with the previous views but also with common intuitions. To talk about concepts means for most people to talk about abstractions; but if concepts are represented by their exemplars, there appears to be no room for abstractions. So we first need some rationale for this seemingly bold move.

Aside from a few extreme cases, the move is nowhere as bold as it sounds because the term *exemplar* is often used ambiguously; it can refer either to a specific instance of a concept or to a subset of that concept. An exemplar of the concept clothing, for example, could be either "your favorite pair of faded blue jeans" or the subset of clothing that corresponds to blue jeans in general. In the latter case, the so-called "exemplar" is of course an abstraction. Hence, even the exemplar view permits abstractions.[1]

A second point is that some models based on the exemplar view do not exclude summary-type information (for example, the context model of Medin and Schaffer, 1978). Such models might, for example, represent the information that "all clothing is intended to be worn" (this is summary information), yet at the same time represent exemplars of clothing. The critical claim of such models, though, is that the exemplars usually play the dominant role in categorization, presumably because they are more accessible than the summary information.

These rationales for the exemplar view accentuate the negative—roughly speaking, the view is plausible because its representations are *not* really restricted to specific exemplars. Of course, there are also positive reasons for taking this view. A number of studies in different domains indicate that people frequently use exemplars when making decisions and categorizations. In the experiments of Kahneman and Tversky (1973), for example, it was found that when subjects try to estimate the relative frequencies of occurrence of particular

From chapter 9 in *Concepts: Core Readings*, ed. E. Margolis and S. Laurence (Cambridge, MA: MIT Press, 1981/1999), 207–221. Reprinted with permission.

classes of events, they tend to retrieve a few exemplars from the relevant classes and base their estimates on these exemplars. To illustrate, when asked if there are more four-letter words in English that (1) begin with *k* or (2) have *k* as their third letter, subjects consistently opt for the former alternative (which is incorrect); presumably they do so because they can rapidly generate more exemplars that begin with *k*. In studies of categorization, subjects sometimes decide that a test item is *not* an instance of a target category by retrieving a counterexample; for example, subjects base their negative decision to "All birds are eagles" on their rapid retrieval of the exemplar "robins" (Holyoak and Glass, 1975). And if people use exemplar retrieval to make negative decisions about category membership, they may also use exemplars as positive evidence of category membership (see Collins and Loftus, 1975; Holyoak and Glass, 1975).

The studies mentioned above merely scratch the surface of what is rapidly becoming a substantial body of evidence for the use of exemplars in categorical decisions (see, for example, Walker, 1975; Reber, 1976; Brooks, 1978; Medin and Schaffer, 1978; Kossan, 1978; Reber and Allen, 1978). This body of literature constitutes the best rationale for the exemplar view.

Concept Representations and Categorization Processes

The Critical Assumption

There is probably only one assumption that all proponents of the exemplar view would accept: The representation of a concept consists of separate descriptions of some of its exemplars (either instances or subsets). Figure 12.1 illustrates this assumption. In the figure the concept of bird is represented in terms of some of its exemplars. The exemplars themselves can be represented in different ways, partly depending on whether they are themselves subsets (like robin, bluejay, and sparrow) or instances (the pet canary "Fluffy"). If the exemplar is a subset, its representation can consist either of other exemplars, or of a description of the relevant properties, or both (these possibilities are illustrated in figure 12.1). On the other hand, if the exemplar is an instance, it must

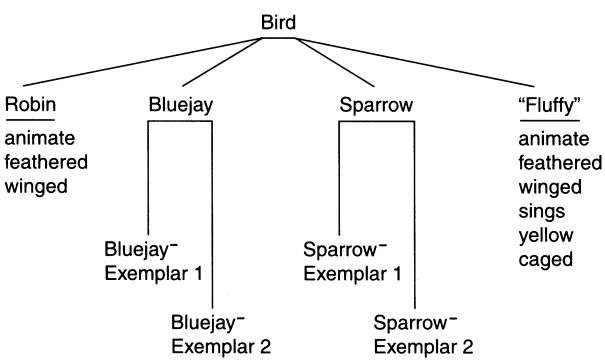


Figure 12.1
An exemplar representation.

be represented by a property description. In short, the representation is explicitly disjunctive, and the properties of a concept are the sum of the exemplar's properties.

This assumption conflicts with that of a summary representation, and it is useful to pinpoint the extent of the conflict. Recall that we use three criteria for a summary representation: it is the result of an abstraction process, it need not correspond to a specific instance, and it is always applied when a question of category membership arises. To what extent is each of these criteria violated by the above assumption? We can best answer this by considering each criterion in turn.

The representation in figure 12.1 shows a clear-cut lack of abstraction in two respects. First, it contains a specific instance, "Fluffy"; second, it contains subsets (for example, robin and bluejay) whose properties overlap enough to permit some amalgamation. Note, however, that the very fact that some exemplars are subsets means that some abstraction has taken place. Thus lack of abstraction is a matter of degree, and our safest conclusion is that exemplar-based representations show a substantially greater lack of abstraction than representations based on the classical or the probabilistic view. This aspect, as we shall see, is the only thing common to all present models based on the exemplar view; so it is the real meat of the critical assumption.

The representation in figure 12.1 also seems at odds with our second criterion, for it contains a component corresponding to a specific instance. Again, the offender is our friend "Fluffy." But if we remove this instance, the representation still qualifies as an exemplar one. That is, some models based on the exemplar view (for example, Medin and Schaffer, 1978) permit representations with no specific instances. Thus, whether or not part of a representation corresponds to an instance is a point on which various exemplar models vary, not a criterion for being an exemplar model.

Finally, there is the summary-representation criterion that the same information is always accessed when category membership is being determined. This issue concerns categorization processes, so the sample representation in figure 12.1 is neutral on this point. Once we consider categorization models based on the exemplar view, it turns out that some violate this criterion (for example, different test items would access different exemplars in the representation in figure 12.1), while others are consistent with the criterion (for example, the entire representation in figure 12.1 would always be accessed when there is a question of birdhood). Again, then, the criterion is really a choice point for various exemplar models.

The Proximity Model as an Extreme Case

We have seen that the critical assumption behind the present view is that the representation lacks abstraction and is "needlessly disjunctive." All exemplar models violate this criterion of a summary representation. Exemplar models differ among themselves, however, with respect to the other two criteria of summary representations; consequently some exemplar models depart from previous views more than others. To appreciate this, it is useful to consider briefly an extreme case of the exemplar view, the *proximity* model (see Reed, 1972). This model violates all three criteria of a summary representation.

In the proximity model each concept is represented by all of its instances that have been encountered. When a novel test item is presented along with a target category, the test item automatically retrieves the item in memory that is most similar to it. The test item will be categorized as an instance of the target concept if and only if the retrieved item is a known instance of that concept. Thus: (1) the concept representation is lacking entirely in abstraction; (2) every exemplar in the representation is realizable as an instance; and (3) the information retrieved in making a decision about a particular concept varies with the test item presented.

Since the proximity model leaves no room at all for abstraction, it conflicts with the intuitions we mentioned earlier. There is another obvious problem with the model. Adults have experienced an enormous number of instances for most natural concepts, and it seems highly implausible that each instance would be a separate part of the representation; the memory load seems too great. For an exemplar model to be plausible, then, there must be some means of restricting the exemplars in the representation. The models that we now consider attempt to do this.

Models of Categorization

Best-Examples Model

Assumptions Though Rosch explicitly disavows a concern with models (1975, 1978), her work—and that of her collaborator, Mervis (1980)—points to a particular kind of categorization model. In the following discussion, we will try to develop it.

In addition to the assumption of exemplar descriptions, the best-examples model assumes that the representation is restricted to exemplars that are typical of the concept—what Rosch often refers to as the *focal instances* (1975). More specifically:

> 1. The exemplars represented are those that share some criterial number of properties with other exemplars of the concept; that is, the exemplars have some criterial family resemblance score. (Since family resemblance is highly correlated with typicality, this amounts to assuming that the exemplars represented meet some criterial level of typicality.)

This assumption raises some questions. First, why leave room for multiple typical exemplars rather than restricting the representation to the single best example? A good reason for not using such a restriction comes directly from data. Inspection of actual family resemblance scores indicates that usually a few instances share the highest score (Rosch and Mervis, 1975; Malt and Smith, 1981). Similarly, inspection of virtually any set of typicality ratings (for example, Rips, Shoben, and Smith, 1973; Rosch, 1975) shows that two or more instances attain comparable maximal ratings. Another reason for permitting multiple best examples is that some superordinate concepts seem to demand them. It is hard to imagine that the concept of animal, for instance, has a single best example; at a minimum, it seems to require best examples of bird, mammal, and fish.

A second question about our best-examples assumption is, How does the learner determine the best exemplars? This question is difficult to answer; all we can do is to mention a few possibilities. At one extreme, the learner might first abstract a summary representation of the concept, then compare this summary to each exemplar, with the closest matches becoming the best exemplars, and finally discard the summary representation. Though this proposal removes any mystery from the determination of best examples, it seems wildly implausible. Why bother with determining best examples when you already have a summary representation? And why ever throw the latter away? A second possibility seems more in keeping with the exemplar view. The learner stores whatever exemplars are first encountered, periodically computes the equivalent of each one's family resemblance score, and maintains only those with high scores. The problem with this method is that it might attribute more computations to the learner than are actually necessary. Empirical data indicate that the initial exemplars encountered tend to have high family resemblance scores; for instance, Anglin's results (1977) indicate that parents tend to teach typical exemplars before atypical ones. This suggests a very simply solution to how best examples are learned—namely, they are taught. The simplicity is misleading, however; for now we need an account of how the teachers determine the best examples. No doubt they too were taught, but this instructional regress must stop somewhere. At some point in this account there must be a computational process like the ones described above.

In any event, given a concept representation that is restricted to the most typical exemplars, we can turn to some processing assumptions that will flesh out the model. These assumptions concern our paradigm case of categorization—an individual must decide whether or not a test item is a member of a target concept. One possible set of assumptions holds that:

> 2a. All exemplars in the concept representation are retrieved and are available for comparison to the test item.
>
> 2b. The test item is judged to be a concept member if and only if it provides a sufficient match to at least one exemplar.

If the matching process for each exemplar is like one of those considered in previous chapters [of Smith and Medin 1981—EM & SL]—for example, exemplars and test item are described by features, and a sufficient match means accumulating a criterial sum of weighted features—then our exemplar-based model is a straightforward extension of models considered earlier. Since few new ideas would arise in fleshing out this proposal, we will adopt an alternative set of processing assumptions.

The alternative is taken from Medin and Schaffer's context model (1978). (Since this is the only exemplar model other than the best-examples model that we will consider, it simplifies matters to use the same set of processing assumptions.) The assumptions of interest are as follows:

> 3a. An entity X is categorized as an instance or subset of concept Y if and only if X retrieves a criterial number of Y's exemplars before retrieving a criterial number of exemplars from any contrasting concept.

3b. The probability that entity X retrieves any specific exemplar is a direct function of the similarity of X and that exemplar.

To illustrate, consider a case where a subject is given a pictured entity (the test item) and asked to decide whether or not it is a bird (the target concept). To keep things simple, let us assume for now that categorization is based on the first exemplar retrieved (the criterial number of exemplars is 1). The presentation of the picture retrieves an item from memory—an exemplar from some concept or other. Only if the retrieved item is a known bird exemplar would one categorize the pictured entity as a bird (this is assumption 3a). The probability that the retrieved item is in fact a bird exemplar increases with the property similarity of the probe to stored exemplars of bird (this is assumption 3b). Clearly, categorization will be accurate to the extent that a test instance is similar to stored exemplars of its appropriate concept and dissimilar to stored exemplars of a contrast concept.

The process described above amounts to an induction based on a single case. Increasing the criterial number of exemplars for categorization simply raises the number of cases the induction is based on. Suppose one would classify the pictured entity as a bird if and only if k bird exemplars are retrieved. Then the only change in the process would be that one might retrieve a sample of n items from memory ($n > k$) and classify the pictured item as a bird if and only if one samples k bird exemplars before sampling k exemplars of another concept. Categorization will be accurate to the extent that a test instance is similar to several stored exemplars of the appropriate concept and dissimilar to stored exemplars of contrasting concepts; these same factors will also govern the speed of categorization, assuming that the sampling process takes time.

Note that processing assumptions 3a and 3b differ from the previous ones (2a and 2b) in that the present assumptions postulate that different information in the concept is accessed for different test items. This is one of the theoretical choice points we mentioned earlier.

One more issue remains: How is the similarity between a test instance and an exemplar determined? The answer depends, of course, on how we describe the properties of representation—as features, dimension values, or templates. In keeping with the spirit of Rosch's ideas (for example, Rosch and Mervis, 1975; Rosch et al., 1976), we will use feature descriptions and assume that the similarity between a test instance and an exemplar is a direct measure of shared features.

Explanations of Empirical Phenomena In this section we will briefly describe how well the model of interest can account for the seven phenomena that troubled the classical view.

Disjunctive concepts Each concept representation is explicitly disjunctive—an item belongs to a concept if it matches this exemplar, *or* that exemplar, and so on.

Unclear cases An item can be an unclear case either because it fails to retrieve a criterion number of exemplars from the relevant concept, or because it is as likely to retrieve a criterion number of exemplars from one concept as from another.

Failure to specify defining features There is no reason why the feature of one exemplar should be a feature of other exemplars; that is, the features need not be necessary ones. And since the concept is disjunctive, there is no need for sufficient features.

Simple typicality effect There are two bases for typicality ratings. First, since the representation is restricted to typical exemplars, a typical test item is more likely to find an exact match in the concept. Second, for cases where a test item is not identical to a stored exemplar, the more typical the test item the greater is its featural similarity to the stored exemplars. Both factors should also play a role in categorization; for example, since typical instances are more similar to the stored exemplars of a concept, they should retrieve the criterial number of exemplars relatively quickly. And the same factors can be used to explain why typical items are named before atypical ones when concept members are being listed. That is, the exemplars comprising the concept representation function as retrieval cues, and the cues themselves should be named first, followed by instances most similar to them. As for why typical exemplars are learned earlier, we have already considered means by which this could come about; for example, the learner may use a kind of family-resemblance computation to decide which exemplars to maintain.

Determinants of typicality The fact that typical instances share more features with other concept members is essentially presupposed by the present model.

Use of nonnecessary features As already noted, there is no requirement that the features of one exemplar be true of all other exemplars.

Nested concepts Figure 12.2 illustrates why some instances (for example, robin) are judged more similar to their immediate than their distance superordinates, while other instances (for example, chicken) manifest the reverse similarity relations. In this illustration robin is one of the represented exemplars for bird, but not for animal. This alone makes it likely that robin is rated more similar to bird than to animal. On the other hand, chicken is a represented exemplar of animal but not of bird, thereby making it likely that chicken is rated as being more similar to animal. In essence, the set of exemplars in a concept may shift with the level of concept.

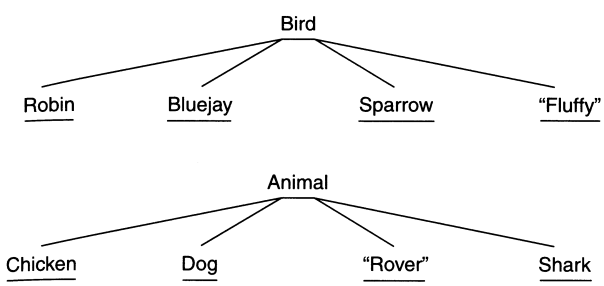


Figure 12.2
Representations that can explain similarity ratings for nested triples.

Category A					Category B				
Instances	Dimension values				Instances	Dimension values			
	EH	ES	NL	MH		EH	ES	NL	MH
1	1	1	1	0	1	1	1	0	0
2	1	0	1	0	2	0	1	1	0
3	1	0	1	1	3	0	0	0	1
4	1	1	0	1	4	0	0	0	0
5	0	1	1	1					

Category A				Category B			
Instances	Dimension values			Instances	Dimension values		
	EH	ES	NL		EH	ES	NL
1′	1	1	1	1′	1	1	0
2′	1	0	1	2′	0	1	1
3′	1	1	0	3′	0	0	0
4′	0	1	1				

Category A	Category B
$.8^a$ high eyes	$.75^a$ low eyes

a = weight associated with dimension value.

Figure 12.3
Representational assumptions of the context model.

Context Model

The context model of Medin and Schaffer (1978) differs from the preceding proposal in two critical respects. One concerns the learning of exemplar representations; the other deals with the computation of similarity in categorization processes. We will consider each issue in turn.

Nature of the Representation To understand the representational assumptions of the context model, we will begin with a simple case. Suppose that subjects in an experiment on artificial concepts have to learn to classify schematic faces into two categories, A and B; the distribution of facial properties for each category is presented abstractly at the top of figure 12.3. Here the relevant properties will be treated as dimensions. They correspond to eye height (EH), eye separation (ES), nose length (NL), and mouth height (MH). Each dimension can take on one of two values, for example, a short or a long nose; these values are depicted by a binary notation in figure 12.3. For example, a nose length of 0 indicates a short nose, a value of 1 signals a long nose. The structure of concepts A and B is presumably that of natural concepts—though A and B lack defining conditions, for each concept there are certain dimension values that tend to occur with its instances. The instances of A, for example, tend to have large noses, while those of B favor small noses.

How, according to the context model, is this information represented by the concept learner? The answer depends on the strategies employed. If our con-

cept learner attends equally to all instances and their dimension values, her final representation should be isomorphic to what is depicted in the top part of figure 12.3—each exemplar would be represented by its set of values. However, if our concept learner selectively attends to some dimensions more than others—say she ignores mouth-height entirely—her representation should be isomorphic to the middle part of figure 12.3. Here instances 2 and 3 of concept A have been collapsed into a single exemplar, and the same is true for instances 3 and 4 of concept B (remember, exemplars can be abstract). This strategy-based abstraction can be even more extensive. To take the extreme case, if our learner attends only to eye height, she will end up with concept representations like those at the bottom of figure 12.3. Here there is no trace of exemplars; instead, the representations are like those in models based on the probabilistic view.

The notion of strategy-based abstraction gives the context model a means of restricting representations to a limited number of exemplars when natural concepts are at issue. (Recall that a plausible exemplar model needs such a restriction.) In particular, suppose that a learner when acquiring a natural concept primarily attends to properties that occur frequently among concept members; then the learner will end up with detailed representations of typical exemplars, which contain the focused properties, but with only incomplete or collapsed representations of atypical exemplars, which do not contain the focused properties. In this way the context model can derive the notion that typical exemplars dominate the representation, rather than assuming this notion outright as is done in the best-examples model. In addition, the context model can assume that in the usual artificial concept study, where there are very few items, each exemplar is fully represented (unless instructions encourage otherwise). Hence in artificial-concept studies, the context model's representations may differ substantially from those assumed by the best-examples model.

Similarity Computations in Categorization The general assumptions about categorization processes in the present model are identical to those in the best-examples model (this is no accident, since we deliberately used the context model's assumptions in developing the best-examples proposal). To reiterate these assumptions:

> 3a. An entity X is categorized as an instance or subset of the concept Y if and only if X retrieves a criterial number of Y's exemplars before retrieving a criterial number of exemplars from any constrasting concept.
>
> 3b. The probability that entity X retrieves any specific exemplar is a direct function of the similarity of X and that exemplar.

There is, however, an important difference between the context model and the previous one with regard to how these assumptions are instantiated. The difference concerns how similarity, the heart of assumption 3b, is computed.

Thus far, whenever we have detailed a similarity computation we have used an *additive* combination. In featural models, the similarity between a test item and a concept representation (whether it is summary or an exemplar) has been some additive combination of the individual feature matches and mismatches.

Sample representations

Category A Instances	Dimension values EH	ES	NL	MH
1	1	1	1	0
2	1	0	1	0
3	1	0	1	1
4	1	1	0	1
5	0	1	1	1

Category B Instances	Dimension values EH	ES	NL	MH
1	1	1	0	0
2	0	1	1	0
3	0	0	0	1
4	0	0	0	0

Sample compulations for Category A exemplars

$$S(A1,A1)^{*} = 1 \cdot 1 \cdot 1 \cdot 1 = 1.0 \qquad S(A2,A1) = 1 \cdot \alpha_{ES} \cdot 1 \cdot 1 = \alpha_{ES}$$

$$S(A1,A2) = 1 \cdot \alpha_{ES} \cdot 1 \cdot 1 = \alpha_{ES} \qquad S(A2,A2) = 1 \cdot 1 \cdot 1 \cdot 1 = 1.0$$

$$S(A1,A3) = 1 \cdot \alpha_{ES} \cdot 1 \cdot \alpha_{MH} = \alpha_{ES} \cdot \alpha_{MH} \qquad S(A2,A3) = 1 \cdot 1 \cdot 1 \cdot \alpha_{MH} = \alpha_{MH}$$

$$S(A1,A4) = 1 \cdot 1 \cdot \alpha_{NL} \cdot \alpha_{MH} = \alpha_{NL} \cdot \alpha_{MH} \qquad S(A2,A4) = 1 \cdot \alpha_{ES} \cdot \alpha_{NL} \cdot \alpha_{MH} = \alpha_{ES} \cdot \alpha_{NL} \cdot \alpha_{MH}$$

$$S(A1,A5) = \alpha_{EH} \cdot 1 \cdot 1 \cdot \alpha_{MH} = \alpha_{EH} \cdot \alpha_{MH} \qquad S(A2,A5) = \alpha_{EH} \cdot \alpha_{ES} \cdot 1 \cdot \alpha_{MH} = \alpha_{EH} \cdot \alpha_{MH}$$

Final categorization

$$P(A1 \subset \text{Category A})^{**} = \frac{1.0 + \alpha_{ES} + \alpha_{ES} \cdot \alpha_{MH} + \alpha_{NL} \cdot \alpha_{MH} + \alpha_{EH} \cdot \alpha_{MH}}{\sum_{X} S(A1,X)}$$

$$P(A2 \subset \text{Category A}) = \frac{\alpha_{ES} + 1.0 + \alpha_{MH} + \alpha_{ES} \cdot \alpha_{NL} \cdot \alpha_{MH} + \alpha_{EH} \cdot \alpha_{ES} \cdot \alpha_{MH}}{\sum_{X} S(A2,X)}$$

*$S(i, j)$ = Similarity between i and j
**$P(i \subset \text{Category A})$ = Probability that i is assigned to Category A

Figure 12.4
How the context model computes similarity.

In dimensional models, similarity between test item and concept representation has been measured by an additive combination of differences on component dimensions. This notion of additivity is rejected by the context model. According to the present model, computing the similarity between test instances and exemplar involves *multiplying* differences along component dimensions.

This process is illustrated in figure 12.4. The top half repeats some representations given in the previous figure. Associated with each dimensional difference is a similarity parameter, α_i, with high values indicating high similarity. Thus α_{NL} is a measure of the similarity between a long and a short nose. Two factors can decrease the size of each parameter, that is, decrease the similarity between the values of a dimension. One factor is the psychophysical difference between the two values of a dimension; the other is the salience of the dimension, which is itself determined by the attentional and strategy considerations

that we discussed earlier. Given a fixed set of parameters, one computes similarity between test item and exemplar by multiplying the four parameters. As examples, the similarity between a test item and exemplar that have different values on every dimension would be $\alpha_{EH} \cdot \alpha_{ES} \cdot \alpha_{NL} \cdot \alpha_{MH}$, while the similarity between a test item and exemplar that have identical values on all dimensions would be $1 \cdot 1 \cdot 1 \cdot 1 = 1$. Some intermediate cases are shown in the middle part of figure 12.4. The bottom part of figure 12.4 shows how these similarity computations between test item and exemplar are cumulated over all relevant exemplars to derive a final categorization of the test item. The probability of assigning a test item to, say, concept A is equal to the sum of the similarities of the test items to all stored exemplars of A, divided by the sum of the similarities of the test item to all stored exemplars of both A and B (this instantiates assumption 3b).

How much hinges on computing similarity by a multiplicative rule rather than by an additive one? Quite a bit, as the two cases illustrated in the middle part of figure 12.4 demonstrate. Following the multiplicative rule, instance 2 should be easier to learn and categorize than instance 1. This essentially reflects the fact that instance 2 is highly similar (that is, differing on only one dimension) to two exemplars of category A (instances 1 and 3) but is not highly similar to any exemplar of concept B; instance 1, on the other hand, is highly similar to only one exemplar in A (instance 2) but to the first two exemplars in B. Had we computed similarity by an additive rule, this prediction would reverse. This can be seen by noting that instance 1 shares an average of more than two values with other exemplars of A, while instance 2 shares an average of exactly two values with other A exemplars. (Both instances share the same average number of values with B exemplars.) These contrasting predictions were tested in a number of artificial-concept experiments by Medin and Schaffer (1978), and the results uniformly supported the multiplicative rule: instance 2 was learned faster and categorized more efficiently than instance 1. In a follow-up study (Medin and Smith, 1981) we found that the superiority of instance 2 held across widely different instructions, including ones that implicitly suggested an additive rule to subjects.

Admittedly, this particular contrast between multiplicative and additive similarity computations is highly specific, and is probably only realizable with artificial materials. Still, it provides some basis for favoring the context model's way of instantiating the exemplar-based processing assumptions over that specified by the best-examples model. Other reasons for favoring the multiplicative rule will be given later in the chapter.

Explanations of Empirical Phenomena There is no need to detail how the context model handles our standard list of phenomena, since these accounts are virtually identical to those given for the best-examples model. Again, the explicitly disjunctive nature of an exemplar-based representation immediately accounts for the existence of disjunctive concepts, the failure to specify defining properties, and the use of non-necessary properties during categorization. And to the extent that the learning strategies posited by the context model eventuate in a representation dominated by typical exemplars, the model would explain typicality effects in the same manner as the best-examples model.

Criticisms of the Exemplar View

Having discussed some of the strengths of the exemplar view, we now consider its weaknesses. We will first take up those difficulties that the present view shares with the probabilistic one; that is, problems in (1) representing all the knowledge in concepts, (2) constraining possible properties, and (3) accounting for context effects. Then we will consider a fourth set of problems—those that are specific to the exemplar view's critical assumption that a concept is represented by a disjunction of exemplars.

Representing More Knowledge in Concepts

To return to our standard example, how can we represent the knowledge that the properties "small" and "sings" tend to be correlated across exemplars of the concept of bird? Note that the solutions we considered in conjunction with the probabilistic view, such as labeling relations between properties, are irrelevant here. For in the present view exemplars tend to be represented separately, so how can we represent something that pertains to all exemplars?

The most promising solution appears to be this: knowledge about a correlation between properties is *computed* from an exemplar-based representation when needed, rather than *prestored* in the representation. We can illustrate with the kind of representation used in the best-examples model. Suppose that the concept of bird is represented by two best examples, one corresponding to robin, the other to eagle. Then one can compute the negative correlation between size and singing ability by noting that the best example that is small (robin) also sings, while the best example that is large (eagle) does not. More generally, to the extent that each best example contains properties that characterize a particular cluster of instances (for example, many of a robin's properties also apply to bluejays and sparrows), then property differences between best examples reflect correlations among properties in the instances at large.

Another kind of additional knowledge that we have routinely been concerned with has to do with variability in properties associated with a concept. Some knowledge of this sort is implicit in any exemplar representation. The different exemplars represented must manifest some differences in their features or dimension values, and one can use these differences to compute estimates of property variability. The problem, though, is that these computations would probably yield smaller estimates of variability than those actually obtained in relevant experiments (Walker, 1975). This would clearly be the case for computations based on best-examples representations, since only a few highly typical exemplars are represented here, and typical exemplars show only limited variation in their properties (see Rosch and Mervis, 1975). The situation seems more promising for the contest model: it is at least compatible with a concept representation containing multiple exemplars, some of which may be atypical, and its representations therefore permit a more realistic computation of property-variability.

Lack of Constraints

There really are two problems involving constraints with the exemplar view: a lack of constraints on the properties associated with any exemplar, and a lack

of constraints on the relations between exemplars included in the same representation. We will treat only the first problem here, saving the second for our discussion of problems specific to the exemplar view.

We start with the obvious. For exemplars corresponding to instances, there is no issue of specifying constraints in the form of necessary or sufficient properties, since we are dealing with individuals. So the following applies only to exemplars that correspond to subsets of a concept, for example, the exemplars "chair" and "table" of the concept "furniture." With regard to the latter kind of exemplar, the problem of unconstrained properties *vis-à-vis* an exemplar is identical to that problem *vis-à-vis* a summary representation. This is so because a subset-exemplar is a summary representation of that subset—there need be no difference between the representation of chair when it is included as one component of an exemplar representation of furniture and when it stands alone as a probabilistic representation. Hence, all our suggestions about how to constrain properties in probabilistic representations apply *mutatis mutandis* to exemplar representations. For the best-examples model, then, there may be a need to specify some necessary features, *or* some sufficient ones, for each exemplar represented in a concept; otherwise we are left with problems such as the exemplar permitting too great a degree of disjunctiveness.

The same, of course, holds for the context model, but here one can naturally incorporate necessary properties via similarity parameters and the multiplicative rule for computing similarity. Specifically, a dimension is a necessary one to the extent that its similarity parameter goes to zero when values on the dimension increasingly differ; and given a near-zero value on one parameter, the multiplication rule ensures that the product of all relevant parameters will also be close to zero. An illustration should be helpful: a creature 90 feet tall might possibly be classified as a human being, but one 9,000 feet tall would hardly be. In the former case, the parameter associated with the height difference between the creature and known human beings would be small but nonzero; in the latter case, the parameter for height difference might be effectively zero, and consequently the overall, multiplicative similarity between creature and human being would be effectively zero regardless of how many other properties they shared. In essence, we have specified a necessary range of values along the height dimension for human beings. To the extent that this is a useful means of capturing property constraints, we have another reason for favoring multiplicative over additive rules in computing similarity.

Context Effects

Thus far little has been done in analyzing context effects of the sort we described in conjunction with the probabilistic view. We will merely point out here what seems to us to be the most natural way for exemplar models to approach context effects.

The basic idea is that prior context raises the probability of retrieving some exemplars in representation. To return to our standard example of "The man lifted the piano," the context preceding "piano" may increase the availability of exemplars of heavy pianos (that is, exemplars whose representations emphasize the property of weight), thereby making it likely that one of them will actually be retrieved when "piano" occurs. This effect of prior context is itself

reducible to similarity consideration; for example, the context in the above sentence is more similar to some piano exemplars than to others. Retrievability is thus still governed by similarity to stored exemplars, and our proposal amounts to increasing the factors that enter into the similarity computation.

The above proposal seems workable to the extent that a representation contains numerous exemplars. If there are only a few exemplars, then many contexts will fail to activate a similar exemplar. To illustrate, consider the sentence "The holiday platter held a large bird," where the context seems to activate a meaning of bird akin to chicken or turkey. If the representation of bird is restricted to a few typical exemplars, like robin and eagle, there is no way the preceding context effect can be accounted for. Since the best-examples model is restricted in just this way, it will have difficulty accounting for many context effects through differential retrievability of exemplars. The context model is less committed to this kind of restriction, and thus may fare better.

Problems Specific to Exemplar Representations

We see two major problems that stem from the assumption that a concept is represented by a disjunction of exemplars. The first concerns the relation between the disjunctions; the second, the learning of summary information. Both can be stated succinctly.

According to the ideas presented thus far, the only relation between the exemplars in a given representation is that they all point to the same concept. But "exemplars that point to the same concept" can be a trait of totally unnatural concepts. For example, let FURDS be the "concept" represented by the exemplars of chair, table, robin, and eagle; again each exemplar points to the same "concept," but this collection of exemplars will not meet anyone's pretheoretical notion of a concept. The point is that the exemplar view has failed to specify principled constraints on the relation between exemplars that can be joined in a representation.

Since any added constraint must deal with the relation between concept exemplars, the constraint must be something that applies to all exemplars. For the concept of furniture, it might be that all the exemplars tend to be found in living spaces, or are likely to be used for some specific purpose. Positing such a constraint therefore amounts to positing something that *summarizes* all exemplars. In short, any added constraint forces a retreat from a pure exemplar representation toward the direction of a summary representation. The retreat, however, need not be total. The summary constraints may be far less accessible than the exemplars themselves (perhaps because the former are less concrete than the latter), and consequently categorization might be based mainly on exemplars. This proposal would leave the currently formulated exemplar models with plenty of explanatory power; it also seems compatible with Medin and Schaffer's statement of the context model (1978), which does not prohibit properties that apply to the entire concept. But whether our proposal is compatible with the spirit behind the best-examples model (that is, the work of Rosch and her colleagues) is at best debatable.

With regard to learning summary information, we are concerned with the situation where someone (say, an adult) tells a concept learner (say, a child) something like "All birds lay eggs." What, according to the exemplar view, is

the learner to do with such information—list it separately with each stored bird exemplar and then throw away the summary information? This seems implausible. What seems more likely is that when one is given summary information, one holds onto it as such. Again, we have a rationale for introducing a bit of a summary representation into exemplar-based models.

Conclusions

With regard to those problems it shares with probabilistic approaches, the exemplar view offers some new ideas about potential solutions. Thus computing property correlations from exemplars that represent different clusters is an interesting alternative to prestoring the correlation, say, by means of a labeled relation. Similarly, accounting for context effects via differential retrieval of exemplars seems a viable alternative to the context-sensitive devices proposed for the probabilistic view. And the context model's multiplicative rule for computing similarity offers a particularly natural way of incorporating necessary properties into representations that can also contain non-necessary ones. But the exemplar view has two unique problems—specifying relations between disjuncts and handling summary-level information—and the solution to these problems seems to require something of a summary representation. This suggests that it would be a useful move to try to integrate the two views.

Note

1. While "your favorite pair of faded blue jeans" is something of an abstraction in that it abstracts over situations, it seems qualitatively less abstract than blue jeans in general, which abstracts over different entities.

References

Anglin, J. M. (1977) *Word, Object and Conceptual Development*. New York: Norton.

Brooks, L. (1978) "Nonanalytic Concept Formation and Memory for Instances." In *Cognition and Categorization*, ed. E. Rosch and B. Lloyd. Hillsdale, NJ: LEA.

Collin, A., and Loftus, E. (1975) "A Spreading Activation Theory of Semantic Processing." *Psychological Review*, 82: 407–28.

Holyoak, K., and Glass, A. (1975) "The Role of Contradictions and Counterexamples in the Rejection of False Sentences." *Journal of Verbal Learning and Verbal Behavior*, 14: 215–39.

Kahneman, D., and Tversky, A. (1973) "On the Psychology of Prediction." *Psychological Review*, 80: 237–51.

Kossan, N. (1978) "Structure and Strategy in Concept Acquisition." Ph.D. Dissertation, Stanford University.

Malt, B., and Smith, E. (1981) "Correlations Structure in Semantic Categories."

Medin, D., and Schafer, M. (1978) "A Context Theory of Classification Learning." *Psychological Review*, 85: 207–38.

Medin, D., and Smith, E. (1981) "Strategies and Classification Learning." *Journal of Experimental Psychology: Human Learning and Memory*, 7(4): 241–253.

Mervis, C. (1980) "Category Structure and the Development of Categorization." In *Theoretical Issues in Reading Comprehension*, ed. R. Spiro, B. Bruce, and W. Brewer. Hillsdale, NJ: LEA.

Reber, A. (1976) "Implicit Learning of Synthetic Languages: The Role of Instructional Set." *Journal of Experimental Psychology: Human Memory and Learning*, 2: 88–94.

Reber, A., and Allen, R. (1978) "Analogical and Abstraction Strategies in Synthetic Grammar Learning: A Functionalist Interpretation." *Cognition*, 6: 189–221.

Reed, S. (1972) "Pattern Recognition and Categorization." *Cognitive Psychology*, 3: 382–407.

Rips, L., Shoben, E., and Smith, E. (1973) "Semantic Distance and the Verification of Semantic Relations." *Journal of Verbal Learning and Verbal Behavior*, 12: 1–20.

Rosch, E. (1975) "Cognitive Representations of Semantic Categories." *Journal of Experimental Psychology: General*, 104: 192–233.

Rosch, E. (1978) "Principles of Categorization." In *Cognition and Categorization*, ed. E. Rosch and B. Lloyd, 27–48. Hillsdale, NJ: LEA.

Rosch, E., and Mervis, C. (1975) "Family Resemblance Studies in the Internal Structure of Categories." *Cognitive Psychology*, 7: 573–605.

Rosch, E., Mervis, C., Gray, W., Johnson, D., and Boyes-Braem, P. (1976) "Basic Objects in Natural Categories." *Cognitive Psychology*, 3: 382–439.

Smith, E., and Medin, D. (1981) *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Walker, J. (1975) "Real World Variability, Reasonableness Judgements, and Memory Representations for Concepts." *Journal of Verbal Learning and Verbal Behavior*, 14: 241–52.

PART VII

Memory

Chapter 13

Memory for Musical Attributes

Daniel J. Levitin

13.1 Introduction

What is memory? As with many concepts in psychology, people have an intuition about what memory is until they are asked to define it. When we try to define memory, and break it up into its components, this becomes a complicated question. We talk about memorizing phone numbers, remembering a smell, remembering the best route to school. We talk about "knowing" that we're allergic to ragweed or that we had a haircut three weeks ago. Is this knowing a form of memory? A panty hose manufacturer boasts that its new fabric has memory. What do all these forms of memory and knowledge have in common? How do they differ? Psychology departments teach whole courses on memory. It is thus impossible to say much of importance about the topic in just a few introductory paragraphs, but what follows is a brief overview of some of the issues in memory research. Then we will discuss memory for musical events in more detail.

13.2 Types of Memory

Psychologists tend to make conceptual distinctions among different types of memory. When we talk about different types of memory, an immediate question that comes to mind is whether these different types are conceptual conveniences, or whether they have an underlying neural basis. There is strong neurological evidence that particular memory systems are indeed localized in separate parts of the brain. The hippocampus and prefrontal cortex, for example, are known to play a role in the encoding and storage of particular forms of memory. However, the computational environment of the brain is massively parallel and widely distributed. It is likely that a number of processes related to memory are located throughout the brain. Further, some of the conceptual labels for memory systems, such as "procedural memory," actually encompass somewhat independent processes that are conveniently categorized together (for pedagogical reasons), but do not necessarily activate a single distinct brain structure. A more detailed discussion of the relation between brain and memory can be found in the book by Larry Squire (1987).

One kind of memory is the immediate sensory memory we experience as image persistence. For example, if you look outside the window on a bright

From chapter 17 in *Music, Cognition, and Computerized Sound*, ed. P. R. Cook (Cambridge, MA: MIT Press, 1999), 209–227. Reprinted with permission.

day and then close your eyes, an afterimage stays on your retina for a few moments. This has been called *iconic memory* by Ulric Neisser (1967). We talk about the auditory equivalent of this as *echoic memory*: for a few moments after hearing a sound (such as a friend's voice) we are usually able to "hear" a trace of that sound in our mind's ear. Richard Atkinson and Richard Shiffrin (1968) referred to these immediate sensory memories as being held in a *sensory buffer*.

When you are holding a thought inside your head—such as what you are about to say next in a conversation, or as you're doing some mental arithmetic—it stands to reason that this requires some type of short-term, or immediate, memory. This kind of memory, the contents of your present consciousness and awareness, has been called "working memory" by Alan Badelley (1990), and is similar to what Atkinson and Shiffrin called short-term memory.

Long-term memory is the kind of memory that most of us think of as memory—the ability to remember things that happened some time ago, or that we learned some time ago (usually more than a few minutes ago, and up to a lifetime ago). For example, you might have stored in long-term memory images from your high school graduation, the sound of a locomotive, the capital of Colorado, or the definition of the word "protractor." (Actually, in the latter case, you might not be able to retrieve a definition of a protractor, but rather a visual image of what one looks like; this is also a form of long-term memory.) One of the important features of long-term memory is its durability. That is, we tend to think of long-term memories as staying with us for perhaps an indefinite period of time. We may not always be able to access them when we want (e.g., when you have somebody's name on the tip of your tongue but can't quite retrieve it), but we have the sense that the memories are "in there." This is in contrast to short-term memories, which decay rapidly without rehearsal, and are not durable unless they somehow are transferred to long-term memory. The sensory memory/short-term memory/long-term memory distinction appears to have validity at the neural level.

Psychologists also talk about different types of long-term memory, but it is not clear that these reflect different neural systems. Rather, they are different kinds of knowledge stored in long-term memory. It can be useful to make these distinctions for conceptual purposes. The psychologist Endel Tulving (1985) makes a distinction between episodic and semantic memory. There is something different between remembering your eighth birthday and remembering the capital of Colorado. Your eighth birthday is an episode that you can remember, one that occupied a specific time and place. There was also a time and place when you first learned the capitol of Colorado, but if you're like most people, you can't remember when you learned it, only the fact itself. Similarly, we remember what words mean, but usually not when and where the learning occurred. This is called *semantic memory*. Remembering how to ride a bicycle or tie your shoe is an example of another type of memory called *procedural memory*.

It is also important to make a distinction between *memory storage* (or encoding) and *memory retrieval*. One of the tricky parts about designing memory experiments is distinguishing between these operations. That is, if a subject cannot recall something, we need to distinguish between an encoding failure

and a retrieval failure. Sometimes using different retrieval cues can bring up memories that seemed previously unreachable. Current memory researchers use a variety of different methods to study remembering, forgetting, storage, and retrieval processes.

13.3 Working Memory Capacity

George Miller (1956) pointed out that working memory has a limited capacity. The number of pieces of information we can juggle in short-term memory at any one time is between 5 and 9, or what he called "7 $\pm$ 2." As a demonstration, try to keep the following series of digits active in memory:

015514804707619

Most people can't keep this many (15) going at once. It is indeed a bit like juggling. But try again, by looking at the numbers when they are rearranged from right to left, as below:

916707408415510

If you're from California, you'll notice that these are the telephone area codes for the northern part of the state. If these are familiar to you, they become grouped—or "chunked," to use Miller's word—and voilà!—suddenly there are only five pieces of information to remember and it is possible to keep them active in working memory. As another example, consider the following string of fifteen letters:

FBICIAUSAATTIBM

If you are able to chunk this into the familiar three-letter abbreviations, the problem is reduced to keeping five chunks in memory, something most people can do easily.

What does chunking have to do with music? People who study ear-training and learn how to transcribe music are probably chunking information. For example, in a typical ear-training assignment, the instructor might play a recording of a four-piece combo: piano, bass, drums, and voice. The student's task is to write down, in real time, the chord changes, bass line, and melody. If you have never done this before, it seems impossible. But with chunking, the problem becomes more tractable. Although the chords on the piano each consist of three, four, five, or more notes, we tend to hear the chord as a chord, not as individual notes. Beyond this, musicians tend to hear not individual chords but chord progressions, or fragments of progressions, such as ii-V-I or I-vi-ii-V. This is analogous to seeing FBI as a chunk and not three individual letters. The chord changes can be parsed this way, and if the listener misses something, the part that is there provides constraints the listener can use to make an educated guess about the part that is missing. You can see the role of contextual constraints in reading. It is not hard to guess what the words below are, even though each is missing a letter:

basso_n cof_ee

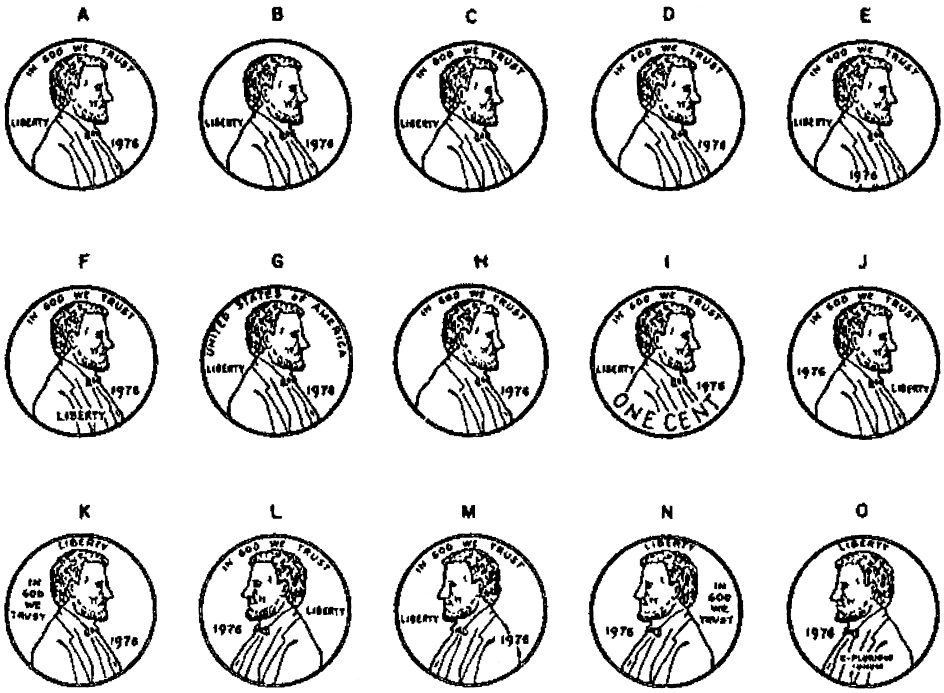


Figure 13.1
Subjects had difficulty identifying the real penny. (Reprinted with permission from Nickerson and Adams, 1979.)

13.4 Remembering and Forgetting Details

A common intuition is that the sole function of memory is to preserve the details of different experiences we've had. But there is a large body of research showing that our memory for details is actually pretty poor. Raymond Nickerson and Marilyn Adams (1979) showed people pictures of different pennies (figure 13.1). Americans see pennies every day, but people in the study could not reliably pick out the accurate picture. Similarly, people tend not to have a very good memory for the exact words of a conversation, but instead remember the "gist" of the conversation. What is the function of memory, then, if not to remember events accurately?

If you think about it, you can see that if we stored and retrieved every detail we encountered every day, we would soon become overloaded with millions of details. When children are first learning language, for example, it is important that they learn to generalize from specific experiences. When a child learns the concept (and the word) "car" as his/her mother points to a car in the street, the child has to somehow disregard the differences among the different cars (the perceptual details) and extract what is common among them. A child who fails to do this, fails to learn the concept of car properly, or to use language properly. That is, the word "car" doesn't apply just to the 1981 Red Honda Accord the child first saw; it applies to objects that share certain properties.

This doesn't necessarily mean the perceptual details are lost: the child may maintain a vivid image of the exact car; but the conceptual system of the brain, along with the memory system, by necessity must integrate details into generalizations. In fact, there is a great deal of evidence that memory does preserve both the details and the "gist" of experiences, and we are usually able to access information at the appropriate level.

13.5 Memory for Music

Objects in the visual world have six perceptual attributes: size, color, location, orientation, luminance, and shape. What do we mean by "object"? This definition has been the subject of heated argument among theorists for many years. I propose that an object is something that maintains its identity across changes (or transformations) in these attributes. In other words, as we move an object through space, it is still the same object. If you were to change the color of your car, it will still be your car. Shape is a tricky attribute, because shape distortions can sometimes, but not always, alter an object's identity. For example, as was shown by William Labov (1973), a cup becomes a bowl if the ratio of its diameter to its height becomes too distorted.

A performance of music contains the following seven perceptual attributes: pitch, rhythm, tempo, contour, timbre, loudness, and spatial location (one might add reverberant environment as an eighth). Technically speaking, pitch and loudness are psychological constructs that relate to the physical properties of frequency and amplitude. The term *contour* refers to the shape of a melody when musical interval size is ignored, and only the pattern of "up" and "down" motion is considered. Each one of these eight attributes can be changed without changing the others. With the exception of contour, and sometimes rhythm, the recognizability of the melody is maintained when each of these attributes is changed. In fact, for many melodies, even the rhythm can be changed to some degree and the melody will still be recognizable (White 1960).

To elaborate further, a melody is an auditory object that maintains its identity under certain transformations, just as a chair maintains its identity under certain transformations, such as being moved to the other side of the room or being turned upside down. A melody can generally retain its identity with transformations along the six dimensions of pitch, tempo, timbre, loudness, spatial location, and reverberant environment; sometimes with changes in rhythm; but rarely with changes in contour. So, for example, if you hear a song played louder than you're accustomed to, you can still identify it. If you hear it at a different tempo, played on a different instrument, or coming from a different location in space, it is still the same melody. Of course, extreme changes in any of these dimensions will render the song unrecognizable; a tempo of one beat per day, or a loudness of 200 dB SPL might stretch the limits of identification.

A specific case of transformation invariance for melodies concerns pitch. The identity of a melody is independent of the actual pitches of the tones played. A melody is defined by the pattern of tones, or the relation of pitches to each other. Thus, when we transpose a melody, it is still recognizable as the same melody. In fact, many melodies do not have a "correct" pitch, they just float

freely in pitch space, starting anywhere one wants them to. "Happy Birthday" is an example of this. Now, you might object to all this and say that Beethoven's String Quartet in F Major ought to be played in F major, and that it loses something when it is transposed. The timbre of the stringed instruments changes with range, and if the piece is played in C major, the overall spectrum of the piece sounds different to the careful listener. But listeners will still recognize the melody because the identity of the melody is independent of pitch.

A number of controlled laboratory experiments have confirmed that people have little trouble recognizing melodies in transposition (Attneave and Olson 1971; Dowling 1978, 1982; Idson and Massaro 1978). Also, at different times and different places, the tuning standard has changed; our present A440 system is arbitrary and was adopted only during the twentieth century. The absolute pitch of the melody's tones is not the most important feature. It is the pattern, or relation of pitches, that is important.

Note the parallel here with our earlier discussion of generalization and abstraction in memory. One of the reasons we are able to recognize melodies is that the memory system has formed an abstract representation of the melody that is pitch-invariant, loudness-invariant, and so on. We take for granted that our memory system is able to perform this important function. Recent evidence suggests that memory retains both the "gist" and the actual details of experience. But what about melodies? Do we retain pitch details, like the absolute pitch information, alongside the abstract representation? This is an interesting question that we will take up in section 13.9, after first reviewing research on memory for contour, lyrics, and failures of musical perception known as *amusias*.

13.6 Contour

Recall that the term *contour* refers to the shape of a melody when musical interval size is ignored, and only the pattern of "up" and "down" motion is considered. At first, the idea of contour being an important attribute of melody seems counterintuitive. Contour is a relatively gross characterization of a song's identity. However, its utility has been shown in laboratory experiments. There is evidence that for melodies we do not know well (such as a melody we have only heard a few times), the contour is remembered better than the actual intervals (Massaro, Kallman, and Kelly 1980). In contrast, the exact interval patterns of familiar melodies are well remembered, and adults can readily notice contour-preserving alterations of the intervallic pattern (Dowling 1994). Infants respond to contour before they respond to melody; that is, infants cannot distinguish between a song and a melodic alteration of that song, so long as contour is preserved. Only as the child matures is he/she able to attend to the melodic information. Some animals show a similar inability to distinguish different alterations of a melody when contour is preserved (Hulse and Page 1988). One explanation of why the contour of a melody might be more readily processed is that it is a more general description of the melody, and it subsumes the interval information. It is only with increasing familiarity, or increasing cognitive abilities, that the intervallic details become perceptually important.

13.7 Lyrics

The memory of ballad singers and tellers of epic poetry has been the focus of a great deal of recent research. On the surface, their memory capacity seems unbelievable for the amount of detail they can readily access. But Wanda Wallace and David Rubin of Duke University have shown that in fact these performers do not need to rely on remembering every detail, because the structures of songs and poems provide multiple constraints for the lyrics (Wallace and Rubin 1988a, 1988b). These constraints are based in part on rhyme, rhythm, alliteration, melodic emphasis, style, and story progression. As an example of lyric constraints, word phrases tend to have a unique stress pattern, such as weak-strong or strong-weak. Similarly, melodic phrases tend to be characterized by strong-weak or weak-strong patterns of accents. Thus, changing a word sequence could alter an entire line's rhythm.

Wallace and Rubin found that from one telling to another, minor alterations in the lyrics occur within these constraints. In a study of eleven singers performing the same ballad on two different occasions, they found that most of the lyric variations conformed to poetic and semantic constraints of the ballad. For example, many lyric changes are to synonyms or other words that do not affect the meaning, rhyme, or rhythm:

(a) "Can't you shovel in a little more coal" becomes

(a′) "Saying shovel in a little more coal"; or

(b) "She cried, 'Bold captain, tell me true'" becomes

(b′) "She cried, 'Brave captain, tell me true.'"

The lyrics and storyline together provide multiple redundant constraints to assist the recall of a passage. For example, even without music, given the first line of the following rock song, the last word of the second line is relatively easy to infer:

"*Well, today a friend told me the sorry tale*
As he stood there trembling and turning———
He said each day's harder to get on the scale."
(From A. Mann, "Jacob Marley's Chain," 1992)

The correct word to end the second line is "pale." Similarly, if one could recall the entire second line except for the word "pale," semantic constraints leave few alternatives. When one adds the contribution of melodic stress patterns, it becomes apparent that our recall of song lyrics is assisted by a number of constraints.

The experimental data corroborate our intuition that the memory representation for lyrics seems to be tied into the memory representation for melody (Serafine, Crowder, and Repp 1984). Further evidence of this comes from a case report of a musician who suffered a stroke caused by blockage of the right cerebral artery. After the stroke, he was able to recognize songs played on the piano if they were associated with words (even though the words weren't being presented to him), but he was unable to recognize songs that were purely instrumentals (Steinke, Cuddy, and Jacobson 1995).

13.8 Amusia

Amusia is the name given to a broad class of mental deficits, involving music perception, that usually appear after brain damage. The deficits include a sharp decrement in an individual's ability to grasp musical relationships in the perception of sounds, or in the ability to perform, read, or write music. Most amusiacs are capable of understanding spoken language, presumably because their neurological impairment spared the speech centers of the brain. However, in many cases amusia accompanies various auditory and speech disorders, such as the aphasias (the name given to various impairments in the production or perception of speech).

The degree to which music and speech rely on common neural mechanisms is not clear. A wealth of cases have shown clear dissociations between impairments in music and in speech, although there may also be individual differences in the way that music is handled by brains. Indeed, in many cases, amusia and aphasia co-occur. There are some separate brain structures, and some shared structures for processing music and speech. For example, Tallal, Miller, and Fitch (1993) found that some children who have trouble learning to speak are unable to process the correct temporal order of sounds. Presumably, if this is a low-level deficit (i.e., a deficit in a brain system shared by music and speech systems), it would also affect the ability to process the order of tones in a melody.

Our current knowledge of the brain's functional architecture is growing rapidly, in part due to advances in neuroimaging techniques. PET (positron-emission tomography), fMRI (functional magnetic resonance imaging), and ERP (event-related potentials) are three such techniques that are allowing neuroscientists to better localize specific brain functions (Posner and Levitin 1997). For example, neuroscientists have demonstrated that there are specific brain anatomies for reading (Posner and Raichle 1994), listening to music (Sergent 1993), mentally practicing one's tennis serve (Roland 1994), calculating numbers (Dehaene 1998), and imagining a friend's face (Kosslyn 1994). Lesions to certain parts of the brain render patients unable to recognize faces (known as prosopagnosia—Bruce 1988; Young and Ellis 1989), although their perception of other objects seems unimpaired. Other lesions cause an inability to read whole words (a type of alexia), although individual letters can still be made out.

Because music performance and perception involve a number of disparate and specialized skills, amusia includes a wide range of deficits. One patient developed an inability to read music note-by-note, but had an intact ability to read whole musical passages. In another case, a musician lost the ability to play the piano (his second instrument) although his ability to play the violin (his first instrument) remained intact. A pianist suffering from aphasia and alexia was unable to read written music or recognize previously familiar melodies; however, her music production abilities were spared, so that she could sing the melody and lyrics to many songs. Following brain damage, an aphasic composer could no longer understand speech but continued to compose without impairment (Luria 1970).

A knowledge of some of the details of brain architecture makes clearer how some of these dissociations can occur. For example, reading music depends a

great deal on the integration of spatial and form perception, because the identity of a musical note is determined both by its form and by its position on the musical staff. An established fact in neuroscience is that form perception and location perception follow different pathways in the visual system (Zeki 1993). It is easy to see how musical alexia (an inability to read musical notes) could arise from damage to either of these two visual pathways, since reading music requires perception of both form and position. It is also easy to see that this damage would not necessarily interfere with other musical skills.

A relatively common dissociation is that found between lyric and melodic production. Oscar Marin (1982) reports the case of an aphasic patient who could sing with normal intonation and rhythm, so long as she wasn't required to sing lyrics. Her ability to join lyrics with melodies was totally impaired.

The neurological syndrome called auditory agnosia is a more general and severe perceptual deficit that usually arises from bilateral damage to the temporal lobes, in particular the auditory cortex (Heschl's area). Patients with auditory agnosia are unable to organize the sounds in the environment, so that speech, animal sounds, bells, and other noises are perceived as a jumbled, uninterpretable stream of noise. A few cases of purely musical agnosia have been described in which patients are unable to organize music into a coherent percept, although their ability to understand speech and nonmusical stimuli remains intact. The extent to which they can understand the "music" of normal speech (known as "prosody") has not been studied thoroughly. For example, are they able to distinguish a question from a statement if the only cue is a rising contour at the end of the sentence? These remain questions open for study.

13.9 Memory for Musical Pitch and Tempo

To what extent do our memories of music retain perceptual details of the music, such as the timbre, pitch, and tempo of songs we have heard? Do we remember all the details of the piece, even details that are not theoretically important? Specifically, since melody is defined by the relation of pitches and rhythms, it would be easy to argue that people do not need to retain the actual pitch and tempo information in order to recognize the song. However, the music theorist Eugene Narmour (1977) argued that listening to music requires processing of both absolute information (schematic reduction) and relative information (irreducible idiostructural), so the question is whether both types of information reach long-term memory.

If people do encode the actual pitches of songs, this would be something like having "perfect pitch" or "absolute pitch" (AP). If you play a tone on the piano for most people and ask them which tone you played, they cannot tell you (unless they watched your hand). The person with AP can reliably tell you "that was a C#." Some APers can even do the reverse: if you name a tone, they can produce it without any external reference, either by singing or by adjusting a variable oscillator. Those with AP have memory for the actual pitches in songs, not just the relative pitches. In fact, most APers become agitated when they hear a song in transposition because it sounds wrong to them.

It has been estimated that AP is rare, occurring in only 1 out of 10,000 people. However, AP studies tend to test only musicians. There is an obvious reason

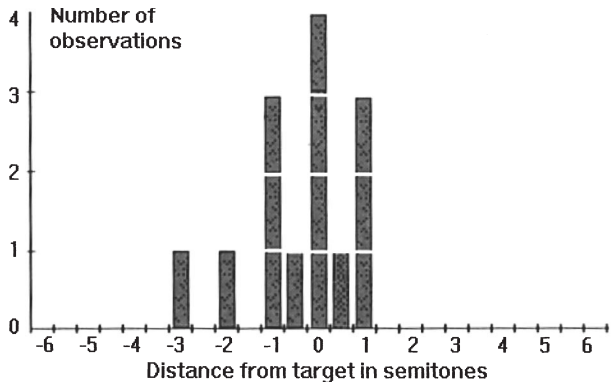


Figure 13.2
Results of pitch memory in non-musicians. The subjects were asked to retain the pitch of a tuning fork in memory for one week.

for this—if you ask most non-musicians to sing an "E-flat," they will not understand. As a term project when I was a student in the Stanford CCRMA psychoacoustics class, I designed a test to determine whether non-musicians could demonstrate AP capabilities. The first test was to determine if non-musicians had an ability to remember pitches over a long period of time—even if they hadn't learned the fancy labels that musicians use. These subjects were given tuning forks, and they were asked to carry the forks around with them for a week, bang them every so often, and to try to memorize the pitch that the forks put out. After a week the tuning fork was taken away, and a week later the subjects were tested on their memory for the tone. Some of them were asked to sing it, and others had to pick it out from three notes played to them. The distribution of the subjects' productions is shown in figure 13.2. Notice that the modal response was perfect memory for the tone, and those who made errors were usually off by only a small amount.

Perhaps, then, absolute musical pitch is an attribute of sound that is encoded in long-term memory. In spite of all the interference—the daily bombardment by different sounds and noises—the subjects were able to keep the pitch of the tuning fork in their heads with great accuracy. A harder test would be to study non-musicians' memory for pitch when that pitch is embedded in a melody. Because melodies are transposition-invariant, the actual pitch information may be discarded once a melody is learned. On the other hand, if somebody hears a melody many times in the same key, we might expect that repeated playings would strengthen the memory trace for the specific pitches.

To test whether people can reproduce the absolute pitch of tones embedded in melodies, I asked subjects to come into the laboratory and sing their favorite rock 'n' roll song from memory (Levitin 1994). The premise was that if they had memorized the actual pitches of the songs, they would reproduce them. It would then be easy to compare the tones they sang with the tones on the original compact disc (CD) version. Rock songs are especially suited to this task because people typically hear them in only one version, and they hear this over and over and over again. Contrast this with "Happy Birthday" or the national

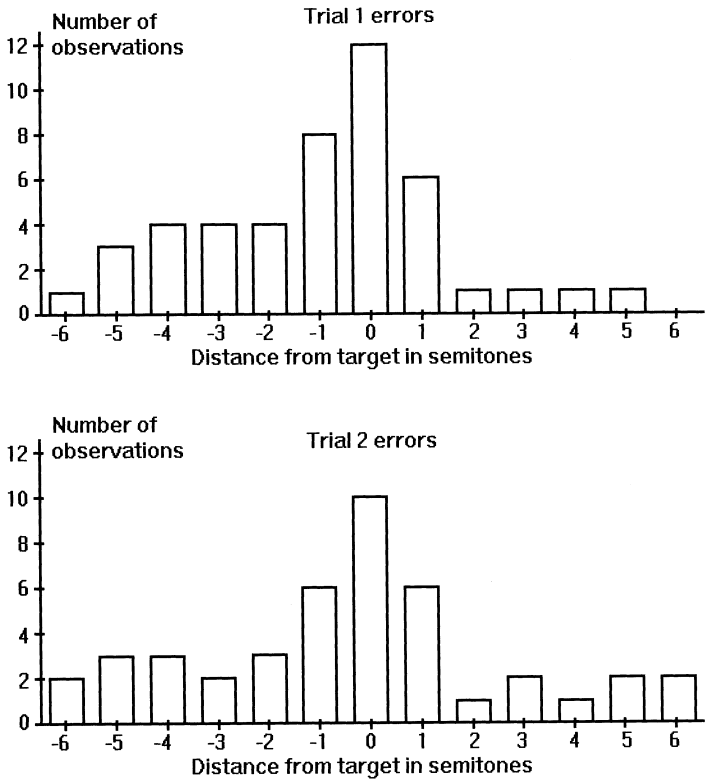


Figure 13.3
Results of pitch memory for the first tone of rock songs. (Upper) Trial 1; (Lower) trial 2.

anthem, which have no objective key standard, and are likely to be sung in a variety of different keys.

The subjects were mostly introductory psychology students (and a few graduate students), they were not specially selected for musical ability or inability, and they didn't know ahead of time that they'd be participating in a music experiment. After they selected a song, they were asked to imagine that it was playing in their heads, and to sing or hum along with it when they were ready.

The subjects could sing as much or as little of the song as they wanted, and they could start wherever they wanted. The first five tones they sang were analyzed, then compared with the five corresponding tones on the CD. There was no difference in accuracy among the five tones or the average of the five tones. Octave errors were ignored (as is customary in absolute pitch research), and how many semitones they were away from the correct tone on the CD was recorded. Thus, the subjects could deviate from the correct pitch by six semitones in either direction.

Figure 13.3 is a plot of the distribution of the subjects' errors. If subjects were no good at this task, their errors would be uniformly distributed among the error categories. In fact, as the top portion of the figure shows, the modal response was to sing the correct pitch. Notice also that the errors cluster around the correct pitch in a mound-shaped distribution. In fact, 67 percent of the

subjects came within two semitones of the correct pitch. The subjects sang a second song, and the findings were essentially the same (lower portion of the figure). Subjects were also consistent across trials, that is, if they were correct on the first song, they were likely to be correct on the second song. From these data, it appears that these nonmusical subjects have something much like absolute pitch. Instead of asking them to "sing a C# or a G," we can ask them to "sing 'Hotel California' or 'Papa Don't Preach,'" and they produce the correct tone. Whether or not they've learned the specialized vocabulary of the musician seems less important than the fact that they have learned to associate a consistent label with a specific tone. This finding has been replicated several times as of this writing (Ashley 1997; Levitin 1996; Wong 1996).

A number of people have wondered if these results might be the product of something other than long-term memory for pitch. If people sing along with their favorite songs, the argument goes, they may have merely developed a "muscle sense" or "kinesthetic sense" from singing the song, and their knowledge of the proper vocal cord tension is driving the results. However, "muscle memory" is a form of long-term memory. All this argument does is specify the subsidiary mechanism in long-term memory that is at work. In addition, it turns out that muscle memory is not very good. W. Dixon Ward and Ed Burns (1978) asked vocalists to sing pitches from memory while being denied auditory feedback (loud white noise in headphones was used to mask the sound of their own voice). The singers were forced to rely solely on muscle memory to produce the requested tones. Their results showed errors as great as a major third, indicating that muscle memory alone cannot account for the precision of performance of the subjects in the sing-your-favorite-rock-song study.

These data support the idea that long-term memory encodes the absolute pitch of songs, even with a group of subjects in whom AP was not thought to exist. This finding also extends Narmour's theory about the two components required for musical perception, showing that both absolute and relative information are retained in long-term memory. A form of *latent* or *residue* absolute pitch is also implied by Fred Lerdahl and Ray Jackendoff's *strong reduction hypothesis* (1983).

Can a song's tempo be accurately encoded as well? The data collected for the pitch study were reanalyzed to test memory for tempo (Levitin and Cook 1996). The subjects weren't explicitly instructed to reproduce tempo during the experimental session, so to the extent that they did, they did so on their own. Tempo would not necessarily have to be explicitly represented in memory, because a melody's identity does not depend on its being heard at exactly the same tempo every time. Because pitch and tempo are separable dimensions (Kubovy 1981), it is possible that one would be preserved in memory and the other would not.

Some interesting properties of song memory are related to the idea of separable dimensions. When we imagine a song in our heads, most of us can easily imagine it in different keys without changing the speed of the song. This is not how a tape recorder works: if you speed up the tape to raise the key, you automatically speed up the tempo as well. Similarly, we can mentally scan a song at various rates without altering the pitch. If you are asked to determine as quickly as possible whether the word "at" appears in "The Star Spangled Ban-

ner," you will probably scan through the lyrics at a rate faster than you normally sing them. This does not necessarily raise your mental representation of the pitch.

In addition, different sections of songs seem to carry "flags" or "markers" that serve as starting points. If you were asked to sing the third verse of "The Twelve Days of Christmas," you might start right on the line: "On the third day of Christmas, my true love gave to me . . ." without having to start from the very beginning. Markers in songs are to some extent idiosyncratic, and depend on what parts of a song are salient, and how well you know the song. Few people are able to jump immediately to the word "at" in "The Star Spangled Banner," but some might be able to start singing it from the phrase "whose broad stripes and bright stars" without having to start from the beginning.

With respect to the other attributes of songs, most people can imagine a song being played loud or soft, being heard in their left ear or right ear or both, being performed inside or outside a large church, and the main melody being carried by various instruments. Most of these things can be imagined even if they have never been experienced before, just as we can imagine a polka-dot elephant, although it's unlikely we've ever seen one.

It is striking to listen to the tapes of non-musical subjects singing, superimposed on the corresponding passage from the CD. They are only singing along with their memory, but it appears that they hear the recording in their head. Enormous amounts of detail appear to be remembered—the subjects reproduce vocal affectations and stylistic nuances, so that it's hard to imagine they could perform any better if they were singing along with the CD.

It wasn't immediately obvious that people would encode tempo with great accuracy, but the data shown in figure 13.4 suggest that they do. As shown in that plot of subject-produced versus actual tempo, 72 percent of the subject's productions were within 8 percent of the correct tempo. How close is 8 percent? Carolyn Drake and Marie-Claire Botte (1993) found that the perceptual threshold for changes in tempo (the just-noticeable difference, or JND) was 6.2–8.8 percent. Thus it appears that people encode tempo information in memory with a high degree of precision.

We have seen that music has a number of different attributes, and that some of these attributes appear to be stored in memory in two forms: a relative encoding of relations and an absolute encoding of sensory features. The precision with which other attributes of musical performances, such as timbre and loudness, are encoded in memory, is the topic of experiments currently under way.

13.10 Summary

The modern view is that memory is distributed throughout various parts of the brain, and that different types of memory engage separate neural structures. Memory for music, just like memory for prose or pictures, probably comprises different cognitive subsystems to encode the various aspects of music. There is a growing consensus that memory serves a dual function: it abstracts general rules from specific experiences, and it preserves to a great degree some of the details of those specific experiences.

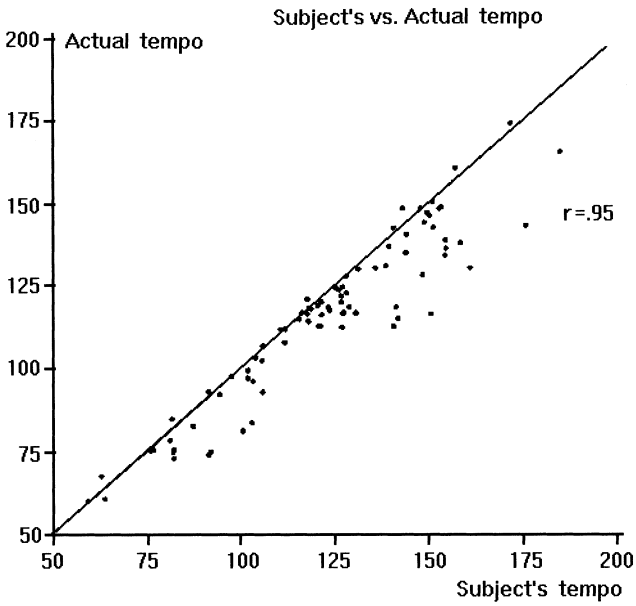


Figure 13.4
Bivariate scatter plot of actual tempo versus produced tempo of rock songs.

Acknowledgments

This chapter benefited greatly from comments by Michael C. Anderson, Gregg DiGirolamo, Gina Gerardi, Lewis R. Goldberg, and Douglas L. Hintzman. I received direct support from a graduate research fellowship from ONR (N-00014-89-J-3186), and indirect support both from CCRMA and from an ONR Grant to M. I. Posner (N-00014-89-3013).

References

Ashley, C. (1997). "Does Pitch Memory Change with Age?" Paper presented at Illinois Junior Academy of Science meeting University of Illinois at Urbana.

Atkinson, R. C., and R. M. Shiffrin. (1968). "Human Memory: A Proposed System and Its Control Processes." In K. W. Spence and J. T. Spence, eds., *The Psychology of Learning and Motivation,* vol. 2, 89–105. New York: Academic Press.

Attneave, F., and Olson, R. K. (1971). "Pitch as a Medium: A New Approach to Psychophysical Scaling." *American Journal of Psychology,* 84, 147–166.

Baddeley, A. (1990). *Human Memory: Theory and Practice*. Boston: Allyn & Bacon.

Bruce, V. (1988). *Recognizing Faces*. Hillsdale, N.J.: Lawrence Erlbaum.

Dehaene, S. (1996). "The Organization of Brain Activations in Number Comparisons: Event Related Potentials and the Additive-Factors Method." *Journal of Cognitive Neuroscience,* 8 (1), 47–68.

Dowling, W. J. (1978). "Scale and Contour: Two Components of a Theory of Memory for Melodies." *Psychological Review,* 85 (4): 341–354.

———. (1982). "Melodic Information Processing and Its Development." In D. Deutsch, ed., *The Psychology of Music*. New York: Academic Press.

———. (1994). "Melodic Contour in Hearing and Remembering Melodies." In R. Aiello and J. A. Sloboda, eds., *Musical Perceptions,* 173–190. New York: Oxford University Press.

Drake, C., and Botte, M.-C. (1993). "Tempo Sensitivity in Auditory Sequences: Evidence for a Multiple-Look Model." *Perception & Psychophysics,* 54 (3): 277–286.

Hulse, S. H., and S. C. Page. (1988). "Toward a Comparative Psychology of Music Perception." *Music Perception,* 5 (4): 427–452.

Huxley, P. (1987). "Double Our Numbers." On the Columbia Records album *Sunny Nights*.

Idson, W. L., and D. W. Massaro. (1978). "A Bidimensional Model of Pitch in the Recognition of Melodies." *Perception and Psychophysics,* 246, 551–565.

Ivry, R. B., and R. E. Hazeltine. (1995). "The Perception and Production of Temporal Intervals Across a Range of Durations: Evidence for a Common Timing Mechanism." *Journal of Experimental Psychology: Human Perception and Performance,* 21 (1): 3–18.

Janata, P. (1995). "ERP Measures Assay the Degree of Expectancy Violation of Harmonic Contexts in Music." *Journal of Cognitive Neuroscience,* 7 (2): 153–164.

Kosslyn, S. (1994). *Image and Brain*. Cambridge, Mass.: MIT Press.

Kubovy, M. (1981). "Integral and Separable Dimensions and the Theory of Indispensable Attributes." In M. Kubovy and J. Pomerantz, eds., *Perceptual Organization*. Hillsdale, N.J.: Lawrence Erlbaum.

Lerdahl, F., and R. Jackendoff. (1983). *A Generative Theory of Tonal Music*. Cambridge, Mass.: MIT Press.

Levitin, D. J. (1994). "Absolute Memory for Musical Pitch: Evidence from the Production of Learned Melodies." *Perception & Psychophysics,* 56 (4): 414–423.

———. (1996). "Mechanisms of Memory for Musical Attributes." Doctoral dissertation, University of Oregon, Eugene, OR. Dissertation Abstracts International, 57(07B), 4755. (University Microfilms No. AAG9638097).

Levitin, D. J., and P. R. Cook. (1996). "Memory for Musical Tempo: Additional Evidence That Auditory Memory Is Absolute." *Perception & Psychophysics,* 58 (6): 927–935.

Loftus, E. (1979). *Eyewitness Testimony*. Cambridge, Mass.: Harvard University Press.

Luria, A. R., Tsvetkova, L. S., and Futer, D. S. (1965). "Aphasia in a Composer." *Journal of Neurological Science,* 2, 288–292.

Mann, A. (1992). "Jacob Marley's Chain," on the Imago Records album *Whatever*.

Marin, O. S. M. (1982). "Neurological Aspects of Music Perception and Performance." In D. Deutsch, ed., *The Psychology of Music*. New York: Academic Press.

Massaro, D. W., Kallman, H. J., and Kelly, J. L. (1980). "The Role of Tone Height, Melodic Contour, and Tone Chroma in Melody Recognition." *Journal of Experimental Psychology: Human Learning and Memory,* 6 (1): 77–90.

Miller, G. A. (1956). "The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review,* 63, 81–97.

Narmour, E. (1977). *Beyond Schenkerism: The Need for Alternatives in Music Analysis*. Chicago: University of Chicago Press.

Neisser, U. (1967). *Cognitive Psychology*. Englewood Cliffs, N.J.: Prentice-Hall.

Nickerson, R. S., and M. J. Adams. (1979). "Long-Term Memory for a Common Object." *Cognitive Psychology,* 11, 287–307.

Pavlov, I. P. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. London: Oxford University Press.

Posner, M. I., and D. J. Levitin. (1997). "Imaging the Future." In R. L. Solso. ed., *Mind and Brain Sciences in the 21st Century,* 91–109. Cambridge, Mass.: MIT Press.

Posner, M. I., and M. E. Raichle. (1994). *Images of Mind*. New York: Scientific American Library.

Roland, P. (1994). *Brain Activation*. New York: Wiley-Liss.

Schacter, D. (1987). "Implicit Memory: History and Current Status." *Journal of Experimental Psychology: Learning, Memory, & Cognition,* 13 (3): 501–518.

Serafine, M. L., R. G. Crowder, and B. Repp. (1984). "Integration of Melody and Text in Memory for Songs." *Cognition,* 16, 285–303.

Sergent, J. (1993). "Mapping the Musician Brain." *Human Brain Mapping,* 1, 20–38.

Squire, L. R. (1987). *Memory and Brain*. New York: Oxford University Press.

Steinke, W. R., L. L., Cuddy, and L. S. Jacobson. (1995). "Evidence for Melodic Processing and Recognition Without Perception of Tonality in an Amusic Subject." Paper presented at Society for Music Perception and Cognition Conference, Berkeley, Calif.

Tallal, P., S. Miller, and R. H. Fitch. (1993). "Neurobiological Basis of Speech: A Case for the Preeminence of Temporal Processing." In P. Tallal, A. M. Galaburda, R. Llinas, and C. von Euler, eds., *Temporal Information Processing in the Nervous System: Special Reference to Dyslexia and Dysphasia,* 27–47. New York: New York Academy of Sciences.

Tulving, E. 1985. "How Many Memory Systems Are There?" *American Psychologist*, 40, 385–398.

Wallace, W. T., and D. C. Rubin. (1988a). "Memory of a Ballad Singer." In M. M. Gruneberg, P. E. Morris, and R. N. Sykes, eds., *Practical Aspects of Memory: Current Research and Issues,* vol. 1, *Memory in Everyday Life*. Chichester, U.K.: Wiley.

Wallace, W. T., and D. C. Rubin. (1988b). "'The Wreck of the Old 97': A Real Event Remembered in Song." In U. Neisser and E. Winograd, eds., *Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory*. New York: Cambridge University Press.

Ward, W. D., and E. M. Burns. (1978). "Singing without Auditory Feedback." *Journal of Research in Singing and Applied Vocal Pedagogy*, 1, 24–44.

White, B. W. (1960). "Recognition of Distorted Melodies." *American Journal of Psychology*, *73*, 100–107.

Wong, S. (1996). "Memory for Musical Pitch in Speakers of a Tonal Language." Undergraduate honors thesis, University of Oregon, Eugene.

Young, A. W., and H. D. Ellis. (1989). *Handbook of Research on Face Processing*. Amsterdam: North Holland.

Zeki, S. (1993). *A Vision of the Brain*. Oxford: Blackwell.

Chapter 14

Memory

R. Kim Guenther

Donald Thompson, a noted expert on memory and a frequent expert witness in legal cases involving eyewitness memories, became a suspect in a case himself when he was found to match a rape victim's description of her rapist. Luckily, Thompson had an airtight alibi—he had been doing an interview on live television, where he was discussing how people can improve their memory for faces. He was cleared when it became apparent that the victim had been watching Thompson on television just prior to the rape and so had confused him with her memory of the actual rapist (this case is described in Schacter, 1996). Indeed, a number of cases have been reported in which eyewitnesses to crimes provided erroneous identifications of perpetrators after they encountered the accused outside the context of the crime (Read, Tollestrup, Hammersley, McFadzen, & Christensen, 1990; Ross, Ceci, Dunning, & Toglia, 1994). Why do people make such mistakes? What accounts for the fallibility of human memory?

In this chapter I will provide an overview of what cognitive psychologists have learned about memory, including how we learn new information, how we recollect previous experiences, and why we sometimes forget important information. I will focus on *explicit memory*, sometimes called *episodic memory*, which is our conscious recollection of personal experiences. In other chapters I will discuss the unconscious influence of past experiences on current thought and behavior and the physiological basis for memory and forgetting.

14.1 Perspectives on Memory

Record-Keeping versus Constructionist Accounts of Memory

I will begin the discussion with the question: What is the principle function of human memory? One possible answer is that memory functions to preserve the past—that it is designed to retain records of previous experiences. Such a perspective has lead to an approach to memory I will label the *record-keeping* approach.

The essential idea of any record-keeping theory is that memory acts as a kind of storage bin in which records of experiences are placed, much as books might be placed in a library. The record keeping theory is really a family of theories that have in common the following principles: (1) Each experience adds a new

From chapter 4 in *Human Cognition* (New York: Prentice-Hall, 1998), 112–156. Reprinted with permission.

record of the experience to the storage bin; consequently the number of records expands over time. Similarly, the number of books stored in a library increases over time. The records actually stored may be more accurately described as interpretations of experiences. (2) Remembering involves searching through a network of memory locations for some particular record, as one might search for a particular book in a library. Once found, the target memory record is "read" or in some sense reexperienced. The search need not be done haphazardly, since the memory records may be connected or organized in such a way as to improve the efficiency of the search. Libraries, for example, organize books by subject matter in order to make finding the books easier. (3) Forgetting is primarily due to search failure caused by the interfering effect of the presence of lots of memory records, just as in a library the huge number of books stored there makes finding any one book difficult. Some versions of the record-keeping theory claim that no memory record is ever really lost. All records of past experiences are potentially recoverable.

The metaphor of record keeping is compelling for several reasons. The word *memory* implies a preserving of the past; we sometimes have vivid and accurate recollections of the past, and nearly all of the artificial memory systems we know about, such as libraries, videotapes, and computers, are record-keeping systems designed to preserve information. Indeed, it is difficult to imagine any other basis for memory. Nevertheless, I will argue in this and other chapters that the record-keeping approach to human memory is a misleading one (Schacter, 1996). Human memory works according to a different set of principles.

An alternative to the record-keeping approach may be called a *constructionist* approach to memory. We know that knowledge from sources outside of the stimulus stream affects the perception of the stimulus. A similar notion plays a role in a constructionist account of memory.

The constructionist account begins with the important insight that human memory is not designed primarily to preserve the past, but to anticipate the future (Morris, 1988). Most constructionist theories are characterized by these principles: (1) Each new experience causes changes in the various cognitive systems that perceive, interpret, respond emotionally, and act on the environment, but no record-by-record account of the experiences that gave rise to those changes is stored anywhere. That is, memory reflects how the cognitive systems have adapted to the environment. Usually this adaptation takes the form of noting regularities in experiences and basing future responses on these regularities. The cognitive systems are also sensitive to unexpected exceptions to the regularities ordinarily observed. (2) Recollection of the past involves a reconstruction of past experiences based on information in the current environment and on the way cognitive processing is currently accomplished. Remembering is a process more akin to fantasizing or planning for the future than searching for and then "reading" memory records, or in any sense reexperiencing the past. The past does not force itself on a passive individual; instead, the individual actively creates some plausible account of her or his past. (3) Forgetting is not due to the presence of other memory records but to the continuous adaptive changes made to the various cognitive systems in response to events.

Let me distinguish between the record-keeping and constructionist approaches with a simple example. Suppose an individual—let's call him Jim—witnessed a robbery in a convenience store. Let's say that the burglar was wearing a black sweatshirt and black jeans, stole money from the cash register, and stole a radio that was lying on the counter. Suppose that after the burglar fled, Jim heard a customer claim that the burglar stole a camera. Later on, when questioned by the police and when testifying in a court of law, Jim must try to recollect as accurately as possible the details of the crime. For example, Jim might be asked: "What was the burglar wearing?" or "What did the burglar steal?"

Any record-keeping theory claims that witnessing the crime caused Jim to store a new record (or records) in his memory system. When later asked to recollect the crime, Jim must first search through his memory records until he finds the record representing the crime, and then try to "read" its contents. If Jim correctly answers questions about the crime, it is because he was able to locate the relevant memory record. If Jim forgets, it is because the presence of so many other memory records made it difficult for him to find the appropriate memory record or because he was unable to access all the details stored in the record.

According to constructionist theories, no record-by-record account of past events is maintained in a storage system. Instead, the cognitive systems for interpreting and acting on experiences change as a function of the event. For example, as a result of the crime experience, Jim might learn to avoid convenience stores and to distrust men who wear black clothes. Jim's cognitive systems function to anticipate possible future events. When Jim is asked questions about the crime, he has no memory records to "read." Instead, he uses the knowledge currently available in his cognitive systems to derive a plausible rendition of the past event. For example, he may use his newly acquired distrust of men in black clothes to deduce that the burglar must have worn black clothes. If Jim forgets, it is because his reconstruction of the past event was inaccurate. For example, he may remember something about a camera, and so reconstruct that he saw the burglar steal a camera when, in fact, the burglar stole a radio.

The main organizing theme of this chapter, then, is the contrast between record-keeping and constructionist accounts of memory. A number of cognitive scientists have noted that this contrast is fundamental to understanding approaches to memory (e.g., Neisser, 1967; Bransford, McCarrell, Franks, & Nitsch, 1977; Rosenfield, 1988; Howes, 1990). Still, probably no contemporary theory of memory entirely embodies the record-keeping theory. Even contemporary theories that may be characterized as predominantly record-keeping also make use of constructionist principles (see Bahrick, 1984; or Hall, 1990). For example, a theory based primarily on record-keeping may claim that people resort to reconstructing the past when they fail to find a relevant memory record. So the record-keeping theory discussed in this chapter serves mainly as a basis of contrast to help make clear how memory does not work. Examples of contemporary theories that primarily (but not exclusively) embody record-keeping principles can be found in Anderson (1983), Anderson and Milson

(1989), Atkinson and Shiffrin (1968), Penfield (1969), and Raaijmakers and Shiffrin (1981). Approaches to memory that may be characterized as predominantly constructionist can be found in Bartlett (1932), Bransford et al. (1977), Loftus (1980, 1982), Neisser (1967, 1984), and Schacter (1996). Constructionist approaches to memory are also implicit in neural net (also known as connectionist or parallel distributed processing) models of memory (e.g., Rumelhart, Hinton & Williams, 1986; Grossberg & Stone, 1986; see Collins & Hay, 1994, for a summary). Raaijmakers and Shiffrin (1992) provide a technical description of various contemporary memory models, while Bolles (1988) provides a nontechnical overview of a constructionist approach to memory written by someone outside the field.

Historical Support for Record-Keeping Theories of Memory

Although I will champion the constructionist theory in this chapter, historically it has been record-keeping metaphors that have dominated thinking about memory (Roediger, 1980). The ancient Greek philosopher Plato, in the Theaetetus dialogue, likened memory to a wax tablet on which experiences leave an impression and likened the process of recollection to trying to capture birds in an aviary. We may not always be able to capture the one we seek. Saint Augustine (A.D. 354–430), an important Christian theologian, and John Locke (1631–1704), a British empiricist famous for his claim that there are no innate ideas, both characterized memory as a storehouse containing records of the past. More recently, cognitive psychologists have used libraries (e.g., Broadbent, 1971), keysort cards (e.g., Brown & McNeill, 1966), tape recorders (e.g., Posner & Warren, 1972), stores (e.g., Atkinson & Shiffrin, 1968), and file systems (e.g., Anderson & Milson, 1989) as metaphors for memory.

The modern era of memory research is usually said to have begun with the publication of Hermann Ebbinghaus's *Uber das Gedachtnis* (*On Memory*) in 1885 (Ebbinghaus, 1885; Hoffman, Bringmann, Bamberg, & Klein, 1986). Ebbinghaus presented himself lists of arbitrarily ordered words or syllables (but not nonsense syllables, as is often claimed) and counted the number of recitations it took him to recall the list perfectly. In some experiments he later attempted to relearn those lists; the reduction in the number of trials to learn the list the second time constituted another, more indirect, measure of memory.

From years of doing these experiments, Ebbinghaus established several important principles of memory. One principle, sometimes known as the Ebbinghaus forgetting curve, is that most forgetting takes place within the first few hours and days of learning (see figure 14.1). After a few days, the rate at which information is lost from memory is very slow and gradual. He also showed that as the number of syllables on a list increased, the number of trials to learn the list increased exponentially. A list of 36 items took him 50 times the number of repetitions to learn as a list of 7 items. Ebbinghaus did not just study arbitrarily ordered lists; he also tried to memorize more meaningful information, specifically various sections of the poem *Don Juan*. He found that he needed only one tenth as many recitations to memorize the poem as he needed to memorize the equivalent number of arbitrarily ordered syllables. Meaningful information is easier to memorize.

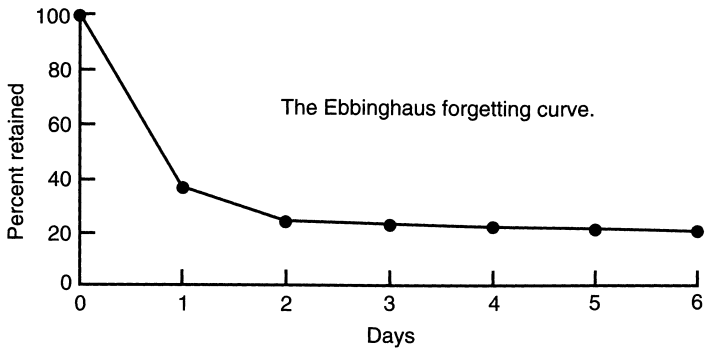


Figure 14.1
The Ebbinghaus forgetting curve.

Ebbinghaus did not spend much time on developing theories about the nature of memory. His primary concern was to demonstrate that human memory is an orderly and measurable phenomenon that can be described with the same precision as biological phenomena. Still, Ebbinghaus's main legacy is his emphasis on memorization of lists of stimuli. Such an emphasis suggests that memory's most important function is to preserve detailed records of past events. Psychologists continue to use experimental methodologies that require subjects to memorize lists of stimuli, such as unrelated words or sentences. Sometimes psychologists make use of Ebbinghaus's relearning paradigm to test memory; more commonly, researchers use *free recall* tests (e.g., "Write down all the words on the lists"), *cued recall* tests (e.g., "What word was paired with *duck* on the list?"), or *recognition* tests (e.g., "Did the word *duck* appear on the list?").

Another development that encouraged the use of record-keeping theories of memory was the invention of the digital computer. Many memory theorists, especially those enamored of the information processing approach to human cognition, have perceived an analogy between how a computer stores information and human memory (e.g., Anderson, 1976, 1983; Winnograd, 1976). Computers store each piece of information by placing records of that information into separate locations, each of which has an address. The memory system in a computer is distinct from the central processing unit (CPU) that actually carries out the manipulation of information. Computers retrieve information either by scanning through the set of locations until the information is found or by going to the address of the memory location and accessing what is stored there. To some theorists, the computer's memory system seems a better metaphor for memory than do passive systems, like libraries. The programs that instruct computers can manipulate and transform stored information, just as we seem to do when we answer questions about and draw inferences from past experiences.

Historical Support for Constructionist Theories of Memory

Although record-keeping metaphors have dominated the history of memory research, there has been a constructionist countertradition. As Brewer (1984) noted, a constructionist conception of memory was the prevalent continental

European view in the 1800s (Ebbinghaus notwithstanding). Sigmund Freud also held to a constructionist approach, writing frequently of how people falsify and remodel their past experiences in the course of trying to recollect them (Freud, 1900/1953; see Erdelyi, 1990). The constructionist approach to memory was introduced to Anglo-American psychology by Frederic Charles Bartlett in his 1932 book *Remembering*. Bartlett was also one of the first to establish a research program investigating the experimental implications of constructionism.

Bartlett's ideas about memory are illustrated in his most famous memory experiments, in which he presented his English subjects an English translation of a Native American folk story called "The War of the Ghosts." The subjects were required to recall the story in as much detail as possible at various time intervals after the story was originally presented to them. The story and one subject's recollection of it are presented in figure 14.2.

"The War of the Ghosts" seems odd to people raised in Western cultures. It includes unfamiliar names, it seems to be missing some critical transitions, and it is based on a ghost cosmology not shared by educated Western people. Bartlett found that his subjects' recollections of the story were incomplete and often distorted. The subjects had trouble remembering the unusual proper names, they invented plausible transitions and, most important, they altered the facts about the ghosts. In fact, many subjects failed to remember anything at all about ghosts. Bartlett claimed that the subjects used their Western cultural knowledge of the nature of stories and other pertinent information to imaginatively reconstruct the story. When relevant cultural knowledge was missing or inappropriate to understanding a story from another culture, the Western subjects' memories were transformed to make their recollections more consistent with their own cultural knowledge. Bartlett's (1932) experiments on memory led him to conclude that remembering is a form of *reconstruction* in which various sources of knowledge are used to infer past experiences.

Another historically influential event in the development of the constructionist tradition was the publication of Ulric Neisser's *Cognitive Psychology* in 1967. In this book Neisser discussed his opposition to the idea that past experiences are somehow preserved and later reactivated when remembered. Instead, Neisser claimed that remembering is like problem solving, a matter of taking existing knowledge and memories of previous reconstructions to create a plausible rendition of some particular past event. Neisser used the analogy of reconstructing a complete dinosaur skeleton from a few bone fragments and knowledge of anatomy. He suggested that "executive routines" guide the process of gathering and interpreting evidence upon which a reconstruction of the past is based. Neisser thought that executive routines were strategies acquired through experience.

Another source of inspiration for a constructionist approach to memory comes from research on the neurophysiology of memory and cognition (see Squire, 1987; Carlson, 1994). Such research has revealed that there is no single place in the brain where past experiences are stored. That is, there does not seem to be anything that corresponds to a storage bin in the brain. Instead, memory reflects changes to neurons involved in perception, language, feeling, movement, and so on. Because each new experience results in altering the strengths of connections among neurons, the brain is constantly "tuning" itself

The War of the Ghosts

One night two young men from Egulac went down to the river to hunt seals, and while they were there it became foggy and calm. Then they heard warcries, and they thought: "Maybe this is a war party." They escaped to the shore, and hid behind a log. Now canoes came up, and they heard the noise of paddles, and saw one canoe coming up to them. There were five men in the canoe, and they said:

"What do you think? We wish to take you along. We are going up the river to make war on the people."

One of the young men said: "I have no arrows."

"Arrows are in the canoe," they said.

"I will not go along. I might be killed. My relatives do not know where I have gone. But you," he said, turning to the other, "may go with them."

So one of the young men went, but the other returned home.

And the warriors went on up the river to a town on the other side of Kalama. The people came down to the water, and they began to fight, and many were killed. But presently the young man heard one of the warriors say: "Quick, let us go home: that Indian has been hit." Now he thought: "Oh, they are ghosts." He did not feel sick, but they said he had been shot.

So the canoes went back to Egulac, and the young man went ashore to his house, and made a fire. And he told everybody and said: "Behold I accompanied the ghosts, and we went to fight. Many of our fellows were killed, and many of those who attacked us were killed. They said I was hit, and I did not feel sick."

He told it all, and then he became quiet. When the sun rose he fell down. Something black came out of his mouth. His face became contorted. The people jumped up and cried.

He was dead.

Subject's Reproduction

Two youths were standing by a river about to start seal-catching, when a boat appeared with five men in it. They were all armed for war.

The youths were at first frightened, but they were asked by the men to come and help them fight some enemies on the other bank. One youth said he could not come as his relations would be anxious about him; the other said he would go, and entered the boat.

In the evening he returned to his hut, and told his friends that he had been in a battle. A great many had been slain, and he had been wounded by an arrow; he had not felt any pain, he said. They told him that he must have been fighting in a battle of ghosts. Then he remembered that it had been queer and he became very excited.

In the morning, however, he became ill, and his friends gathered round; he fell down and his face became very pale. Then he writhed and shrieked and his friends were filled with terror. At last he became calm. Something hard and black came out of his mouth, and he lay contorted and dead.

Figure 14.2
The text of "The War of the Ghosts" and one subject's reproduction of it. From Bartlett, 1932.

in response to experiences. But it has no neural tissue dedicated only to storing a record of each experience.

14.2 Retaining Experiences in Memory

What is it that is retained in our cognitive system as a result of having experiences? The essential idea of a record-keeping theory is that a record of each experience is put into a kind of storage bin. Such records may take a variety of forms, including abstract descriptions or interpretations of events (see Anderson, 1983), lists of items and contextual information (see Raaijmakers & Shiffrin, 1981) or images of the perceptual qualities of events (see Paivio, 1971).

In contrast, the essential idea of a constructionist approach is that the various cognitive systems (e.g., the visual system, the language system) are changed by experiences, but no record-by-record accounts of the experiences are stored anywhere. Instead, the cognitive system is designed to extract the unchanging elements or patterns from experience and to note deviations from enduring patterns.

A Constructionist Account of Retention

To get a somewhat more precise sense of how a constructionist theory explains what is retained from experience, consider this simple example: remembering what you ate for dinner last Thursday night. Research on the effects of diet on health frequently relies on people's memory of what they have eaten. Is memory for food consumption reliable?

In general, research suggests that accurate recall of food items consumed declines to about 55% a week after the consumption (DeAngelis, 1988). The longer the retention interval, the poorer the memory for specific food items consumed (Smith, Jobe, & Mingay, 1991). Over time, people rely more on their generic knowledge of their own dieting behaviors than on a precise memory of any given meal (Smith et al., 1991). In some cases, knowledge of one's own dieting may distort memory. In one study, women on a low-fat diet remembered fewer of the snack items they had eaten the day before than did women on normal or high-fat diets (Fries, Green, & Bowen, 1995). People also tend to underestimate in their memories how much food they have eaten (Fries et al., 1995).

The constructionist account of memory for past meals would go something like the following (see figure 14.3). You have in your cognitive system concepts and ideas about food and food consumption. These include concepts such as iced tea, spaghetti, and entrees as well as ideas such as that snack foods are high in fat content and desserts are served at the end of a meal. The constructionist theory emphasizes that experiences change the strengths of the connections among these ideas and concepts.

To illustrate, suppose that on one night you have spaghetti for an entree and iced tea for a beverage, on the second night you have lamb chops and iced tea, and on the third night you have fried chicken and iced tea. On each night, then, the connections between the ideas of dinner and entree, between the ideas of dinner and beverage, and between the ideas of beverage and iced tea will all be strengthened. These strong connections represent the enduring pattern in the

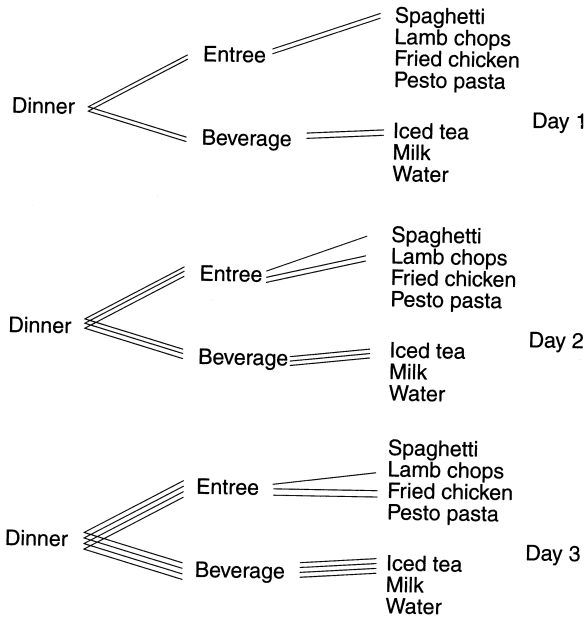


Figure 14.3
Depiction of a constructionist account of memory for three dinners. The more lines that connect one concept to another, the more likely the connections between those concepts will be remembered.

dinner event. On the other hand, your cognitive system will not consistently strengthen the connection between the idea of an entree and the ideas that represent any particular entree (e.g., spaghetti), because the entrees change nightly. For instance, on the second night the connection between entree and lamb chops will be strengthened while the previously established connection between entree and spaghetti will weaken.

If you are later asked what you had for dinner on the first night, the strong connections between the dinner and entree ideas, between the dinner and beverage ideas, and between the beverage and iced tea ideas mean that you will reconstruct that you had some kind of an entree and iced tea. The connections between the idea of entree and any particular entree, such as spaghetti, will be relatively weak; consequently you will not be able to reconstruct as reliably which entree you had the first night. Instead, you may reconstruct only that you had an entree. Note that these reconstructions are accomplished without retrieving an actual record of each night's dinner. Other facts about food consumption may also influence your memory. If you are on a low-fat diet, for example, you may use your knowledge of fat content to deduce that you did not eat potato chips with your meal. In a later section of this chapter I will discuss in more detail how ideas and beliefs affect recollection. Although my example is greatly simplified, it at least illustrates how the cognitive system extracts the invariants of dinner experiences and uses them to form a plausible reconstruction of past dinner experiences.

Constructionist theory, then, predicts that people will not be able to remember very well the constantly changing details of events, such as the particular

entree for any given dinner. Similarly, people might not be able to remember very well such things as what color shirt they wore on any given night out on the town or exactly where in the lot they parked their car on any given excursion to the beach. But it should be easy for people to remember the *invariants* or enduring patterns of events, such as always drinking a beverage with dinner, always wearing a casual shirt to the night club, or always parking in the cheaper lot at the beach.

Record-keeping theories, like constructionist theories, would also predict that accurate memory for any one event is likely to decline as more records are stored (see, for example, Anderson, 1976). But without embellishment, record-keeping theories have no ready way to explain why memory should be strong for the enduring patterns of experience. At the very least, a record-keeping theory would have to postulate the existence of another cognitive mechanism designed only to extract patterns from experiences. That is, it is not a natural consequence of keeping records that enduring patterns are extracted from those records. The advantage of constructionist theory is that it postulates that the creation of memories and the extraction of patterns from experience are accomplished by the same mechanism; namely, the altering of connection strengths among the concepts and ideas that constitute knowledge.

Evidence for the Constructionist Account of Retention

Empirical Evidence That Memory Preserves Patterns but Not Details of Experiences

A nice example of the principle that memory preserves the enduring patterns and themes but not the changing elements in events comes from the testimony of John Dean, a key figure in the Watergate scandal of the early 1970s (Neisser, 1981). John Dean had been President Nixon's attorney and testified against him in a highly publicized Senate hearing on the Watergate break-in. Dean tried to recollect the details of meetings, including who participated, what was said, and when the meeting took place. Dean's memory seemed quite remarkable (and damaging to Nixon); he was able to supply many details that other members of Nixon's administration claimed to be unable to recall.

It was discovered later that all meetings in the Oval Office had been tape recorded, so that many of Dean's recollections could be compared with the actual transcripts of those meetings. It turns out that Dean was often inaccurate about details of the meetings but was accurate in his recollection of the general tenor of a number of the meetings; namely, that Nixon and other high-ranking members of his administration knew about the Watergate break-in and tried to cover it up. What distinguished Dean's testimony from that of the others was that Dean decided to tell the truth about the coverup. Dean's memory was not especially accurate about those elements that were always changing, like the details of conversations or which participants were at particular meetings, but his memory was quite accurate about the sorts of topics and issues that endured across many meetings.

Many memory experiments also make the point that our memories permit easier recall of enduring patterns than of details of specific experiences (e.g., Bartlett, 1932; Bransford & Franks, 1971; Thorndyke & Hayes-Roth, 1979). For example, participants in a weekly seminar on math were asked to recall the

names of the other participants who had attended the last meeting of the seminar (Freeman, Romney, & Freeman, 1987). The subjects were not able to recall very accurately; about half of their responses were errors. The errors were revealing, however. Sometimes subjects mistakenly excluded someone who had attended the last meeting, but usually the excluded person had not regularly attended the seminar. And sometimes subjects mistakenly included someone who had missed the last meeting, but usually the included person had attended most of the other meetings. The errors suggest that the subjects had extracted the general pattern of attendance from their experiences in the seminar and had used that pattern, reasonably enough, to reconstruct who had attended the last meeting.

Memory for patterns is also reflected in the tendency for people to remember the gist but not the details of their experiences. Research has shown that subjects will forget the exact wording of any given sentence in a passage after reading only a few more sentences, but will usually be able to remember the meaning of the sentence (Sachs, 1967; for similar results with pictures, see Gernsbacher, 1985). Research has also shown that after studying a text or a set of pictures, people will tend to believe mistakenly that a sentence or picture was explicitly in the set of information they studied, when, in fact, it was only implied by the information (e.g., Bransford, Barclay, & Franks, 1972; Harris & Monaco, 1978; Maki, 1989; Sulin & Dooling, 1974; Thorndyke, 1976). For example, if a passage describes an event in which a long-haired customer sat in a barber's chair and later left the barbershop with short hair, a subject who had read that passage may mistakenly believe that the passage also contained a sentence describing the barber cutting the man's hair. The reason for the mistake is that the implicit information is likely to be consistent with the passage's essential themes, which would form the basis of the reconstruction of the details of the passage.

That memory is better for the patterns or invariants than for the ever-changing details of experiences is what enables memory to be adaptive, to anticipate the future. It is the invariants of experience that we are likely to encounter in future events, so a cognitive system that readily notices such patterns will be better prepared to respond to new experiences.

Good memory for the patterns or invariants of experience stands in contrast to our extremely poor memory for the details of the majority of experiences. Consider—can you describe in detail what you were doing around 3:00 P.M. on May 15th two years ago? Do you remember what the topic of conversation was when you first met your next-door neighbor? Or what your boss was wearing when you first met him or her? Or the first 10 sentences of this chapter? You see the point. What is especially remarkable about our memories is the almost complete lack of detail they provide about the majority of our past experiences! And it is easy to demonstrate experimentally that people do not remember very much about long-past experiences. For example, people have trouble remembering their infant-rearing practices, such as whether they fed their infants on demand (Robbins, 1963), their formerly held opinions on important political issues, such as whether they supported busing to equalize education (Goethals & Reckman, 1973); whether they voted in any given election (Parry & Crossley, 1950); and what they had to eat for dinner six weeks ago (Smith et al., 1991).

Accurate Memory A possible objection to the constructionist theory is the observation that people can sometimes remember past events accurately. A record-keeping theory of memory claims that accurate memory occurs when a person successfully locates a memory record. How can the constructionist theory account for accurate recollections? And, one might also object, what about people who have extraordinarily accurate memories, who seem to have a memory system that works like a videotape machine?

Constructionist theory implies that there are three circumstances in which memory is likely to be accurate. First, as I have already discussed, constructionist theory predicts that repetitious events, like always having iced tea with dinner, should be well remembered, because they promote the creation of strong connections among elements. A high probability, therefore, exists that at least some of the relevant connections created by the repetitive event will remain stable over time and so permit the accurate reconstruction of that event. Research shows that information that is repeated is more easily remembered than information that is presented only once (e.g., Jacoby, 1978; Greeno, 1964). To be fair, record-keeping theories also predict that repetition improves memory, because repetition would increase the number of records of that event, making any one record easier to find.

Second, constructionist theory predicts that recent events, such as what one ate for breakfast this morning, should be well remembered, because the strength of the connections among elements representing recent events would not yet be weakened by subsequent events. Researchers since Ebbinghaus have observed that recently experienced events are usually the easiest to remember (Ebbinghaus, 1885; Wickelgren, 1972).

Record-keeping theories need a modification to predict that recent events are better remembered. The modification is that recent events are stored in a more accessible manner or location. One way to visualize that is to imagine that events are stored in a push-down stack (Anderson & Bower, 1973). Recent events are first placed at the top of the stack but are gradually pushed further down into the stack by the continuous storage of even more recent events. The retrieval mechanism would begin its search at the top of the stack.

Third, constructionist theory predicts that unusual or distinctive events should be well remembered because they promote the creation of connections among elements that would not likely be reconfigured by future events. Consider an unusual event such as becoming nauseated after eating lamb chops. The connection between the feeling of nausea and the idea of lamb chops is not likely to be diminished by subsequent dinner experiences, because lamb chops would not ordinarily become associated with other ill feelings nor would nausea become associated with other entrees. Any subsequent activation of the lamb chops idea, then, is also likely to activate the feeling of nausea, permitting accurate memory for that experience of nausea.

Record-keeping theories could also predict that distinctive events are better remembered. One way to do so is to imagine that events are stored in locations that reflect the attributes of the event. Memories of happy experiences might be stored in one place, memories of car repair experiences might be stored in another place, and so on. A distinctive event has a collection of attributes that is different from other events and so would be stored in an uncluttered place in

the memory system. It is easier to find a memory record in an uncluttered space than in a cluttered space, just as it would be easier to find *The Joy of Nausea* in a library that had only one book on the topic of nausea than in a library that carried hundreds of books on nausea.

That distinctive events are readily remembered has been well established by research (see Schmidt, 1991, for a review). In one experiment that required subjects to recall words from a list, the subjects were better able to remember that an animal name appeared on the list if the animal name was embedded in a list of names of countries than if the same animal name was embedded in a list of other animal names (Schmidt, 1985). This finding is an example of the *Von Restorff effect*, after the psychologist who first discovered it (Von Restorff, 1933). In another experiment, subjects were given photographs of human faces and were asked to judge the distinctiveness of each face. When later asked to recognize which faces they had previously studied, the subjects more accurately recognized the faces they rated as distinctive than the faces they rated as common (Cohen & Carr, 1975). At least some research shows that events associated with strong emotions, which are presumably distinctive, are better remembered than emotionally more neutral events (e.g., Waters & Leeper, 1936; Holmes, 1972).

Best-selling books on how to improve memory (e.g., Lorayne and Lucas, 1974) encourage the use of bizarre imagery to improve the memorability of verbal information, such as names of people. Bizarre images presumably make information more distinctive. But does the use of bizarre imagery really improve memory? The answer seems to be a qualified yes.

The standard experimental paradigm investigating the role of imagery in memory requires subjects to memorize word pairs (e.g., *chicken–cigar*) by making various kinds of images of the words. The results have shown that when people create bizarre images to connect the words (e.g., a chicken smoking a cigar), they will later recall more of the words than when they create common images (e.g., a chicken pecking a cigar) to connect the words (for a review, see Einstein, McDaniel, & Lackey, 1989). However, the advantage of bizarre over common images usually occurs only when the same person is required to make bizarre images for some of the words on the to-be-remembered list and ordinary images for the rest of the words on the list. When subjects are required to make bizarre images for all the words on the list, then the individual images are not as distinctive, and there is no longer an advantage of bizarre images over common images. Research also suggests that the superiority of the bizarre image technique is greater if the memory test is done days after studying the list (Webber & Marshall, 1978). When the delay between forming the images and recalling the words is only a few minutes, memory for the words is at least as good using the common image technique.

That distinctive events are memorable is also revealed in memory for real-life experiences. Erickson and Jemison (1991) had students record one event from their lives each day for 12 weeks, and 5 months later take several memory tests on the events. They found that the more memorable events tended to be the distinctive ones—that is, the ones rated atypical, infrequent, or surprising. They also found that positive events were more memorable, possibly because positive events are likely to be thought about and discussed frequently.

When we have accurate memories of long-past events, these events are almost always remarkable—that is, distinctive—in some way. For example, I vividly remember a championship Little League baseball game in which I got five hits and scored the winning run (a newspaper account verifies that my memory is accurate). However, about all I remember from the many other Little League games in which I played is that I was good at throwing and catching but not so good at hitting.

Psychologists have studied memory of remarkable experiences by asking people what they were doing on the occasion of some historically significant event like the assassination of John F. Kennedy (Brown & Kulik, 1977; Pillemer, 1984). Usually people can describe what they were doing in great detail, although ordinarily the psychologist is unable to check the accuracy of the person's account. Memory for a remarkable event, sometimes called a *flashbulb memory*, is vivid (McCloskey, Wible, & Cohen, 1988) because the event is distinctive and because people talk about and think about the event much more frequently than about other, more mundane, experiences.

It should be noted, though, that memory for what one was doing at the time of a historically significant event is frequently wrong (McCloskey et al., 1988; Neisser & Harsch, 1991). For example, Neisser and Harsch (1991) asked students on the day after the Challenger disaster how they heard about the disaster and asked them again 3 years later. On the test conducted 3 years after the disaster, one third of the subjects gave inaccurate accounts, although they were confident that their accounts were accurate.

Brain Stimulation and Accurate Memory Sometimes memory researchers cite data that seem to indicate, as the record-keeping theory would have it, that human memory does contain records of nearly all past experiences, although it might ordinarily be hard to retrieve most of those records. Some of the most compelling data comes from the research of a brain surgeon named Wilder Penfield, who removed small portions of cortical tissue in order to prevent the spread of seizures in epileptic patients (Penfield & Jasper, 1954; Penfield & Perot, 1963). Ordinarily such patients are awake during the operation, because the cortex is impervious to pain. Penfield needed to electrically stimulate various portions of the cerebral cortex, in order to locate accurately the epileptic site. When he did so, some of the patients described vivid recollections of mostly trivial past experiences. Penfield reasoned that the cortex must therefore keep a record of all past experiences and that forgetting must be due to retrieval failure.

After Penfield began to publish his findings, some psychologists questioned his interpretations (Loftus & Loftus, 1980; Squire, 1987). First of all, only about 3% of Penfield's patients ever reported remembering past experiences in response to electrical stimulation. Furthermore, for those patients who did, the evidence suggested that they were not accurately recalling an actual experience but unintentionally fabricating one. One patient, for example, reported having a memory of playing at a lumberyard, but it turned out the patient had never been to the lumberyard. Another patient claimed to remember being born.

Recognition and Accurate Memory Another kind of data sometimes cited to support the claim that the brain stores records of virtually all experiences, any

one of which is potentially retrievable, comes from research on recognition memory. In some recognition experiments, subjects are shown thousands of detailed pictures, such as magazine advertisements, and weeks to months later are given a recognition test in which they must discriminate the *old* pictures from *new* ones (e.g., Standing, 1973). In one of these experiments, subjects' recognition accuracy was 87% after one week (Shepard, 1967), while in another experiment recognition accuracy was 63% after a year (chance performance would be 50%) (Standing, Conezio, & Haber, 1970).

However, it also possible to design such experiments so that a person's recognition accuracy is not much better than chance, only minutes after viewing pictures (Goldstein & Chance, 1970). Critical to performance in recognition experiments is the similarity between the *old* stimuli and the *new* stimuli used as foils (Dale & Baddeley, 1962; Pezdek et al., 1988). When *old* and *new* pictures closely resemble one another, recognition accuracy is poor. But when the *old* and *new* pictures are dissimilar, subjects need not remember very much about a set of pictures to distinguish between *old* and *new* ones. Note that pictures of advertisements used in the high-accuracy memory experiments are relatively dissimilar from one another.

Still, the high percentage of correct responses in some recognition experiments does make the important point that we have much better memory for our experiences than we might ordinarily think. How good our memory seems to be for any given event depends critically on how we are tested. As I will discuss later, performance is usually better on recognition than on recall tests and is better the more cues there are in the environment to prompt memory. But it would be a mistake to assume that if a more sensitive test improves memory scores, then all experiences must be stored in, and therefore potentially retrievable from, memory.

Autobiographical Memory Another kind of finding sometimes used to support the notion that nearly all experiences are potentially retrievable comes from individuals who have for years kept records of details of important autobiographical experiences and later tried to recall some of those details (Linton, 1978; R. T. White, 1982, 1989). These individuals seem to remember something about nearly all the events they recorded.

Typical of this research is a study done by Willem Wagenaar (Wagenaar, 1986). Each day for six years Wagenaar selected an event or two and recorded what happened, who he was with when it happened, the date it happened, and where it happened. He tested his memory for an event by reading some details about the event (e.g., "I went to a church in Milano") and trying to recall other details (e.g., "I went to see Leonardo da Vinci's *Last Supper* on September 10, 1983"). He found that even years afterwards he was able to recall at least one detail of about 80% of the events he recorded.

Does his research contradict the constructionist theory that predicts forgetting of most events? I think not. First of all, Wagenaar deliberately selected salient, distinctive events to record; he avoided mundane events. The constructionist theory predicts good memory for distinctive events. It is interesting to note that after about one year, Wagenaar was able to recall accurately slightly less than 50% of the details of even these distinctive events. Further-

more, Wagenaar had no way to control for talking or thinking about the events later on; consequently, many of these events were likely recycled many times through his cognitive systems. Also, he was often able to make plausible guesses about what happened. For example, given the cue "I went to a church in Milano" he may have been able to guess the approximate date by just remembering that his trip to Italy took place during the first two weeks of September in 1983. Finally, Wagenaar had no "foils"—events that could plausibly have happened to him but did not—to see if he could accurately discriminate between real events and foils. In fact, research demonstrates that people have a hard time distinguishing between actually experienced events and plausible foils in their recollections about important autobiographical experiences (Barclay & Wellman, 1986).

In short, research on autobiographical memory does not prove that we have accurate and detailed memory for nearly all of our experiences. It suggests that we can remember, or at least infer, some of the details of our most distinctive experiences.

"Photographic" Memory? But what about individuals who seem to have something akin to a photographic or videotape memory in which all experiences are accurately remembered? Wouldn't the existence of these people contradict the constructionist approach to memory? Incidentally, I do not intend for the notion of photographic memory to imply that the individual has only an especially good memory for visual information. Instead, "photographic" is meant to be a metaphor for extraordinary memory for all kinds of information.

A few extensive investigations of such rarely encountered individuals have been carried out. Probably the best-known memory expert was S. V. Shereshevskii, usually referred to as *S*. *S* grew up around the turn of the century in Latvia and was a Moscow newspaper reporter when his editor noticed his exceptional memory. The editor recommended that *S* have his memory evaluated at the local university; there he met Aleksandr Luria, a great Russian psychologist.

Luria studied *S* over a period of about 30 years (Luria, 1968). Luria verified that *S*'s memory was quite extraordinary. For example, *S* was able to repeat back a series of 70 randomly selected numbers in order after hearing them only once. As another example, he was able to recall lists of arbitrary and randomly ordered words 15 years after Luria presented the words to him. *S* claimed that he formed vivid and detailed images of every stimulus he was asked to remember and often associated the images with images of familiar locations, like Gorky Street in Moscow. He would later retrieve the words from memory by taking a mental "walk," noticing the images associated with the landmarks. This *mnemonic technique* (i.e., a strategy for memorizing) is called the *method of loci*, and can be used effectively by anyone trying to memorize a list of stimuli (Groninger, 1971). Techniques like the method of loci improve memory for several reasons, one of which is that they help make information more distinctive.

S made use of other mnemonic techniques, as well. He seemed to have the exceedingly rare ability, known as *synesthesia*, to conjure up vivid images of light, color, taste, and touch in association with almost any sound. These images also helped him remember new information. For a time, *S* found work as a

memory expert on stage. People would call out words or numbers for him to remember and he would try to recall them exactly. Interestingly, though, *S* sometimes needed to develop new mnemonic techniques to overcome occasional errors in memory and so improve his act. For example, he had difficulty remembering names and faces. If *S* had a photographic memory, he would have been able to memorize accurately any kind of information presented to him. His extraordinary memory, then, was not a result of possessing anything analogous to a photographic mind, but was rather a result of having an appropriate mnemonic strategy. Tragically, *S* ended his life in a Russian asylum for the mentally ill.

Some people, called *eidetic imagers*, seem to have an extraordinary ability to remember visual details of pictures. Eidetic imagers report that, after viewing a picture, they see an image of the picture localized in front of them and that the visual details disappear part by part. While they remember many more visual details of a picture than would the ordinary person, often the accuracy of their reports is far from perfect (Haber & Haber, 1988; see Searleman & Herrmann, 1994).

The all-time champion eidetic imager was an artist known as Elizabeth. Her most remarkable achievement had to do with superimposing two random-dot patterns to see a three-dimensional image. In one experiment (Stromeyer & Psotka, 1970), she was first presented with a 10,000-random-dot pattern to her right eye for 1 minute. The first pattern was then removed for 10 seconds and a second 10,000-random-dot pattern was presented to her left eye. She was instructed to superimpose her memory of the image of the first pattern onto the second. The patterns were designed so that when superimposed and examined through both eyes, a three-dimensional figure (e.g., a square floating in space) would appear. It was impossible to determine the three-dimensional image from either pattern alone, however. Elizabeth was able to superimpose a memory of the first pattern onto the second pattern and thus accurately identify the three-dimensional image. In fact, in one case, she was able to hold a 1,000,000-random-dot pattern in memory for 4 hours and then superimpose her memory of that pattern onto a second 1,000,000-random-dot pattern to identify successfully the three-dimensional image! It is possible to see the three-dimensional figure in the superimposed patterns even when one of the patterns is significantly blurred, although the blurring will also make the edges of the three-dimensional image more rounded. So Elizabeth need not have remembered the exact position of all of the dots to accomplish seeing the three-dimensional figure, although she claimed that the edges of her three-dimensional image were sharp and not rounded.

No one else has yet been found who can come close to Elizabeth's visual memory; indeed, some people are skeptical of her feats (see Searleman & Herrmann, 1994). As far as I know, Elizabeth was not tested for memory of anything other than visual information. It remains unclear, then, whether she had an outstanding all-purpose memory or an extraordinary memory for only visual information.

Another remarkable memorizer is Rajan Mahadevan, who has a phenomenal memory for numbers. He is able to recite the first 31,811 digits of pi from memory (I'm lucky if I can remember the first four digits!). In a series of

experiments comparing his memory to that of college students, Rajan Mahadevan dramatically outperformed the students on any memory test involving numbers (Thompson, Cowan, Frieman, Mahadevan, & Vogel, 1991). For example, he recalled 43 randomly ordered digits presented to him once, while the college students recalled an average of only about 7 digits. Rajan Mahadevan claims that he does not use imagery to help him remember numbers but instead uses a rather vaguely described mnemonic system whereby numbers are associated with numerical locations in a series. It does not seem that he has anything analogous to a videotape or photographic memory, however. His recall for nonnumerical information, such as word lists or meaningful stories, was about equal to that of the average college student. For example, he recalled an average of about 41 ideas from several previously read Native American folk tales similar to "The War of the Ghosts," while the college students recalled about 47 ideas on average from the same stories.

A reasonable conclusion, then, is that individuals like *S* and Rajan Mahadevan make use of mnemonic devices that others could use to help make information more memorable (Ericsson & Polson, 1988; Hunt & Love, 1972). While the memorizing skill of these mnemonists can seem phenomenal, it is clear that their memories do not work like a videotape recorder; otherwise they would be able to remember the details of any and all of their experiences. Instead, their memory is good for classes of information in which they are experts (Elizabeth was a skilled artist) or for which they have learned mnemonic memorizing strategies. The Hollywood version of the person with a "photographic" mind probably does not exist.

The Assimilation Principle

Making information distinctive or associating information with distinctive images and ideas can promote better memory of that information. Such techniques may be called learning strategies. What other learning strategies help make information memorable? Another useful learning strategy is based on the principle that memory for an event will be improved to the extent that the event can be assimilated into something that already exists in memory (Stein & Bransford, 1979; Stein, Littlefield, Bransford, & Persampieri, 1984). This principle is called the *assimilation principle*.

Assimilation means that new information is incorporated into relevant preexisting knowledge useful for interpreting the new information. For example, a passage describing the nature of electricity would be more memorable if the passage reminded readers of their knowledge of rivers. The passage would not be as memorable if it did not remind readers of relevant knowledge, nor would it be as memorable if it reminded readers of irrelevant knowledge, such as their knowledge of baseball. The constructionist theory explains the assimilation principle this way: When new information is assimilated into relevant preexisting knowledge, there is widespread activation of the cognitive system for interpreting an event and an increase in the number and strength of the connections among elements of that cognitive system. Reconstruction of the event is improved to the extent that strong connections among elements in that cognitive system can be found.

Experimental Support for Assimilation A variety of research supports the assimilation principle. One kind of support comes from experiments that show that people remember more new information if that information is within their area of expertise than if the new information is outside their area of expertise (Bellezza & Buck, 1988; Chiesl, Spilich, & Voss, 1979; Morris, 1988). For instance, experienced bartenders remember better than do novices their customers' drink orders (Beach, 1988). Football experts can remember more about descriptions of fictitious football games than nonexperts (Bellezza and Buck, 1988). Chess experts will remember the positions of chess pieces on a chessboard better than chess novices, provided the pieces are arranged in a way consistent with the rules of chess. If the chess pieces are randomly arranged, however, the chess expert can remember their locations no better than the novice (Chase & Simon, 1973).

Sometimes when people must learn new material, like the material in this book, they have a hard time figuring out what general patterns or principles are implied by the material and so are unable to associate the material with the appropriate elements in their cognitive systems. Any aids that help people find such principles in the material will improve memory. If subjects are required to memorize a list of words, they will remember more of them if the words in the list are grouped according to categories, like animal names, than if the words are presented in a random order (Bower, Clark, Lesgold, & Winzenz, 1969; Mandler, 1979). Subjects given titles that clarify the meaning of otherwise obscure pictures or passages remember more than subjects not given titles (Bransford & Johnson, 1972). When subjects read technical or scientific passages, the subjects first given guides to help them associate the information with familiar ideas (e.g., electrical current is like a river) or help them see the relationships among key ideas in the text will later be able to recall more of the text than subjects not first given the guides (Dean & Kulhavy, 1981; Brooks & Dansereau, 1983; Lorch & Lorch, 1985). Most of the advantage for subjects receiving the guides is in remembering the conceptual information and not the technical detail (Mayer, 1980; Mayer & Bromage, 1980).

Levels of Processing and the Assimilation Principle Another manifestation of the assimilation principle is found in investigations of what is usually called *levels of processing* (Craik & Lockhart, 1972; Koriat & Melkman, 1987). This research establishes that when people think about the meaning of information, they remember more of it than when they think about the physical properties or when they merely try to rote memorize the information. Elaborating on the meaning is a more effective learning strategy than is rote memorizing.

In one example of research on levels of processing, subjects studied a list of words by making judgments about each word, and later recalled the words. Subjects recalled more words for which they had been asked to judge "How pleasant is the word?" than words for which they had been asked to judge "Does the word contain the letter *e*?" (Hyde & Jenkins, 1975; Parkin, 1984). Subjects who studied a list of words by elaborating each word into complete sentences (called elaborative rehearsal) later recalled more of the words than subjects who only rote memorized the words (called maintenance rehearsal) (Bjork, 1975; Bobrow & Bower, 1969).

The advantage of processing for meaning is not limited to verbal information. Subjects were better at recognizing pictures of faces if they previously thought about whether each face seemed friendly than if they previously thought about whether each face had a big nose (Smith & Winograd, 1978) and if they assessed faces for honesty rather than for the sex of the face (Sporer, 1991). In general, thinking about the meaning of a stimulus or elaborating on the stimulus is likely to permit the stimulus to be assimilated by a greater portion of a cognitive system, and so create more possibilities for reconstructing a memory of the stimulus later on. Elaboration may also help make information more distinctive (Craik & Lockhart, 1986; Winnograd, 1981).

Processing the meaning of a stimulus improves memory only when that processing connects the stimulus to relevant knowledge. For instance, asking a person whether a shirt is a type of clothing enhances memory for the word *shirt*, as opposed to the case where the person is asked whether the word *shirt* contains more vowels than consonants. However, asking a person whether a shirt is a type of insect does not promote very good memory for *shirt* (Craik & Tulving, 1975). In the latter case, answering the question does not encourage the person to connect *shirt* with knowledge of shirts (see Schacter, 1996).

Levels of processing research has been used to challenge the duplex model of short-term memory (see Klatzky, 1980). There is an important qualification to the general finding that thinking deeply about information promotes better memory than does thinking in a shallow manner about the information. The qualification is that it depends on how memory is tested. If the memory testing procedure matches the manner in which information is originally learned, then memory for that information is better than if there is a mismatch.

An example comes from a study by Morris, Bransford, and Franks (1977). Subjects were required to decide for each of a group of words whether the word could have a particular semantic property (e.g., "Does a train have a silver engine?") or whether the word rhymes with another word (e.g., "Does *train* rhyme with *rain*?"). The semantic task was the "deep" task and the rhyming task was the "shallow" task. Later, some subjects were given a standard recognition task in which they had to pick out the target word from a list of distractors. Subjects who had made the semantic judgment did better on the recognition task than did subjects who had made the rhyming judgment. But other subjects were given a very different test of memory in which they had to pick out from a list of words which word rhymed with one of the words previously studied. Now it was the subjects who had originally made the rhyming judgments who did better. This finding, usually called *transfer appropriate processing*, is discussed again later in this chapter.

Individual Differences in Memory

Why does one person have a better memory than another person? Record-keeping theories, especially those that liken human memory to the memories of computers or libraries, imply that there is an all-purpose memory system for storing every kind of experience. According to the record-keeping theory, the reason some people have better memories than others is that some people have more efficient mechanisms for storing or retrieving records. Even Plato talked about some people having a purer kind of wax tablet for storing experiences.

Constructionist theories, on the other hand, imply that there is no all-purpose memory system. Memory is instead a byproduct of changes to the various components of cognition that underlie perception, language, emotions, and so on. From the perspective of the constructionist approach, there are no storage and retrieval mechanisms whose efficiency varies from person to person. Instead, people vary with respect to how much they know about various domains of knowledge. According to constructionist theory, the main reason some people have better memories than others is that some people have more expertise in the domain of knowledge sampled by the test of memory. For example, a baseball expert can use the knowledge that runners on second base often score after a single to reconstruct that the home team scored a run in the previous inning. However, baseball knowledge would not help the baseball expert remember, say, a passage about climate in South America.

The constructionist theory claims, then, that the best predictor of how well a person remembers new information in some domain, such as baseball, is how much knowledge the person already possesses about that domain. General intellectual skills, especially skill at memorizing lists of information unrelated to the domain, should not predict individual differences in memory for information within some domain. If, instead, memory is an all-purpose system, it would follow that performance on tests of memory and on general intellectual skills would readily predict memory for new information.

The research supports the constructionist theory's explanation of individual differences in memory. Good memory for information within some domain is primarily a function of expertise in that domain and not a function of any general intellectual skill. Schneider, Korkel, and Weinert (1987) and Walker (1987) found that subjects who scored low on a test of general aptitude but happened to know a lot about baseball recalled more facts about a fictitious baseball game than did subjects who scored high on the general aptitude test but knew very little about baseball, and recalled as many facts as did high-aptitude subjects who knew a lot about baseball. Kuhara-Kojima and Hatano (1991) found that knowledge about music, but not performance on a test of memory for unrelated words, predicted how many new facts subjects recalled from a passage about music.

Merely possessing domain knowledge does not guarantee better memory for new information in that domain, however. DeMarie-Dreblow (1991) taught people about birds but found that the newly acquired bird knowledge did not help subjects recall a list of bird names any better than subjects not given the knowledge about birds. The knowledge has to be well-learned, and people need practice using the knowledge in the context of reconstructing a memory for the new information (Pressley & Van Meter, 1994).

For instance, Pressley and Brewster (1990) taught their Canadian subjects new facts about Canadian provinces. Some subjects were given prior knowledge in the form of pictures of some prominent setting in the province. By itself, this prior knowledge did not help subjects remember the new facts all that much better than the subjects not given the prior knowledge. Other subjects were given imagery instructions for which the subjects were to imagine the fact occurring in a setting unique to the province referred to by the new fact. Imagery instructions also did not help subjects all that much. However, subjects

given both the prior knowledge and the imagery instructions did recall substantially more new facts than did subjects who did not have both the prior knowledge and the techniques (i.e., imagery) for using that knowledge to learn and remember new information.

The better predictor of memory for novel information, then, is a person's degree of expertise in that domain (provided the person knows how to use the knowledge for learning and remembering) and not the person's general intellectual level or memorizing ability for unrelated information. The main practical implication is that people develop good memory, not to the extent that they become better memorizers, but to the extent that they develop expertise in domains for which it is important to remember details accurately.

By way of summarizing this section, let me suggest how a student can make use of the material I have discussed. Suppose you must study this chapter on memory in preparation for an exam, and so are required to learn a lot of factual details. What can you do to make the chapter more memorable? Just repeatedly reading the facts will not in itself enhance your memory for this chapter very much. Instead, you must first look for the themes and patterns that serve to organize the material presented in the chapter. For example, the chapter presents two points of view about memory, the record-keeping theory and the constructionist theory, and argues that the constructionist theory is superior. You must then try to understand these themes by relating them to what you already know. You might note that the record-keeping theory is similar to how books are stored in and retrieved from a library. You should then attempt to figure out for each piece of information how it makes a distinctive contribution to the thesis. You might ask what unique insight each experiment makes concerning the predictions of the constructionist theory. Finally, and to anticipate the next section, you should practice studying the material in a way similar to the way you are going to be tested. If you know that the test will be an essay test, then write out answers to essay questions. Remember, human memory is designed to anticipate the future, not recapitulate the past.

14.3 Recollecting the Past

So far I have focused on how cognitive systems change as a result of experiences. Now I wish to change the focus to the cognitive processes responsible for recollecting a past event. What is a good model of recollection?

Record-Keeping and Constructionist Models of Recollecting the Past

The record-keeping approach claims that recollecting the past means searching through a storehouse of records of past events until the target record is retrieved. Finding or "reading" the memory record is like reexperiencing the past event. The search process is thought to be guided by information in the current environment that acts as a sort of address for the location of the target record. The search through the records need not be haphazard, because the records may be organized, much the way books in a library are organized by content.

The constructionist approach to memory claims that recollecting the past is essentially a process of reconstructing the past from information in the current

environment and from the connections serving the various cognitive systems. Recollection typically involves making plausible guesses about what probably happened. Recollection is an active process, akin to fantasizing or speculating about the future, whereby people recreate or infer their past rather than reexperience it. Another way to put it is that people learn reconstruction strategies that enable them to deduce past events. Loftus (1982) provides a discussion of some of the various types of reconstruction strategies.

To illustrate, suppose a person returns to the scene of a car accident and tries to recall the details of the accident, which occurred several days earlier. Returning to the intersection is likely to activate the same elements of the cognitive system involved in originally perceiving the accident; consequently some perceptual details necessary to reconstruct the accident will become available (e.g., cars move quickly through the intersection). Thoughts about a car accident may also activate knowledge of how cars work (e.g., brakes often squeak when a driver tries to stop a rapidly moving vehicle). Such knowledge may then become a basis for reconstructing the accident. Information that was provided to the person after the accident occurred may also be activated and inserted into the reconstruction of the accident (e.g., a friend at the scene of the accident later claimed to have seen a van cut in front of the car). The confluence of activated elements constitutes the memory of the accident (e.g., a van cut in front of a fast-moving car, which tried to stop, causing its brakes to squeal). The memory may appear to the person to be vivid and accurate, yet some details may be in error (e.g., perhaps the van never cut in front of the car).

Reconstructing the Past

An important implication of reconstruction is that when people try to recollect a past event, what they will remember about that event will depend on what they currently know or believe to be true about their lives. Errors in recollecting events will not be haphazard, but will instead reflect knowledge and beliefs. So researchers interested in demonstrating reconstruction often vary a person's current knowledge and show that the person's recollection of some past event will be distorted as a consequence (Dooling & Christiaansen, 1977; Hanawalt & Demarest, 1939; Snyder & Uranowitz, 1978; Spiro, 1977).

A nice demonstration of reconstruction is provided by Spiro (1977). In his experiment, subjects read a passage about a couple. Bob and Margie, who were engaged to be married. Bob was reluctant to tell Margie that he did not want to have children, but, by the end of the story, finally confronted Margie with his wishes. In one version of the story, Margie told Bob that she wanted children very badly. Afterwards, the subjects were told either that Bob and Margie are now happily married or that the engagement had been broken off. Days to weeks later, subjects returned and tried to recall the details of the story. Subjects who were told that the engagement had been broken off tended to recall accurately that Bob and Margie disagreed sharply about having children. In some cases they even exaggerated the disagreement. But the subjects told that Bob and Margie were now happily married tended to recall that the disagreement was much less severe than was actually depicted in the story. And the longer the time between reading the story and recalling it, the more likely these subjects distorted the story so as to resolve the inconsistency between the

disagreement and the subsequent marriage. Furthermore, subjects who incorrectly recalled minimal disagreement between Bob and Margie were every bit as confident of their mistaken recollections as they were of their accurate recollections about other aspects of the story.

These results make sense if we assume that subjects did not activate a memory record of the story, but instead used their belief that successful engagements require agreement about whether to have children, in order to reconstruct the story. If Bob and Margie are still married, then it would have seemed that any disagreement about children must not have been very serious.

An intriguing implication of a reconstructionist approach to memory is that it ought to be possible to create false memories—that is, memories of events that never happened. Some researchers have suggested that some memories of sexual abuse are actually false memories created by psychotherapeutic practices that encourage clients to interpret certain psychological symptoms as evidence of past abuse.

Eyewitness Memory and Reconstruction Reconstruction has been studied extensively in the context of eyewitness memory. A variety of research has shown that eyewitnesses tend to distort their memories of crimes and accidents based on information they receive after the crime or accident.

For example, eyewitness memory research demonstrates what is called *photo bias*. In one experiment on photo bias, subjects were first shown a film of a crime and were later presented photographs of suspects. Later still the subjects were required to pick the actual perpetrator out of a lineup. What happened was that subjects tend to be biased towards identifying as the perpetrator any suspect whose photograph they had recently seen, even when the person was innocent of the crime (Brown, Deffenbacher, & Sturgill, 1977). Apparently, when the subjects viewed the lineup, they recognized that they had seen one of the suspects before, and erroneously assumed that it must be because the suspect was the criminal.

Elizabeth Loftus, one of the most influential advocates of a reconstructionst approach to memory, has conducted a variety of experiments in which subjects are shown a film of an accident and are later asked questions about the film (Loftus, 1979; Loftus, Miller, & Burns, 1978; Loftus & Loftus, 1980; Loftus & Palmer, 1974). In one experiment, she asked one group of subjects leading questions like "Did another car pass the red Datsun while it was stopped at the stop sign?" when, in fact, the Datsun was stopped at a yield sign. (This particular experiment was conducted before Datsun changed its name to Nissan.) When questioned again about the film, these subjects were much more likely to claim they saw the Datsun stop at a stop sign than another group of subjects not initially asked the misleading question. In some cases, memory was tested by showing subjects two slides, a slide of a Datsun stopped at a stop sign and a slide of the Datsun stopped at a yield sign. Most of the misled subjects selected the slide displaying a stop sign, even when the misled subjects were offered a substantial reward ($25) for remembering accurately. Incidentally, this experimental paradigm usually contains a whole set of questions about various details of the accident or crime. I am illustrating the paradigm with only one of the questions that might be used. At any rate, the subjects were presumably

using the information implied by the question to reconstruct the details of the accident. If the question falsely implied that the car stopped at a stop sign, then subjects reconstructed a stop sign in their recollections of the accident.

Exactly what would such a reconstruction be based on? One possibility is that mentioning a stop sign effectively erased or somehow undermined the connection between the accident and the yield sign and replaced it with a connection between the accident and the stop sign (Loftus & Loftus, 1980). There is another possibility, though. Maybe subjects do remember that the film contained, say, a yield sign and that the subsequent question mentioned a stop sign. But when given the choice between a yield and stop sign, the subjects figure that the experimenter wants them to say that they saw a stop sign in the film (otherwise, why would the experimenter ask the question?). In other words, maybe subjects' memories are just fine in this paradigm; maybe they are just responding to the demands characteristic of the experiment; maybe this research is not supportive of a construction approach to memory (McCloskey & Zaragoza, 1985).

To see if the question about the stop sign really erased the information about the yield sign (or, more generally, if misinformation erases previously acquired information), McCloskey and Zaragoza (1985) devised a somewhat different experimental paradigm (this paradigm, the Loftus paradigm, and a couple of other paradigms that I discuss below are all illustrated in figure 14.4). Subjects first saw a film that contained details like the yield sign, and then read a text that contained misinformation, such as a description of a stop sign, and then were asked to decide if the original film contained a yield sign or, say, a caution sign. Again, the actual paradigm includes several pieces of information, and not just information about traffic signs. If the misinformation really wiped out memory of the yield sign, subjects should be just as likely to choose the yield sign as to choose the caution sign. Instead, subjects overwhelmingly selected the correct alternative, the yield sign in this case.

So does this mean that the misinformation has no effect on eyewitness memory at all? No. In other research (see Lindsay, 1993), subjects were first shown a film (or slide show) of a crime or accident, then read a text that contained some misleading information, and then were asked of each piece of information whether the information was presented in the film, in the text, in both places, or in neither place. The memory test in this case (see figure 14.4) asks subjects the source of the information. The idea is to see whether source memory is worse for a detail in the film when there is misinformation in the text than when there is not misinformation in the text.

The main finding is that *source memory* tends to be good (i.e., subjects correctly remember that the stop sign was in the text and not in the film) when it is easy for subjects to discriminate between the experience of seeing the film and reading the text (e.g., Zaragoza & Lane, 1994; see Lindsay, 1993). One way to make the discrimination easy is to present the film on one day, wait until the next day to present the text, and then immediately follow the text with the memory test. Source memory tends to be poor (i.e., subjects think that the stop sign was in the film) when it is hard for subjects to discriminate between the experience of seeing the film and the experience of reading the text. For example, a way to make the discrimination hard is to present the text right after the

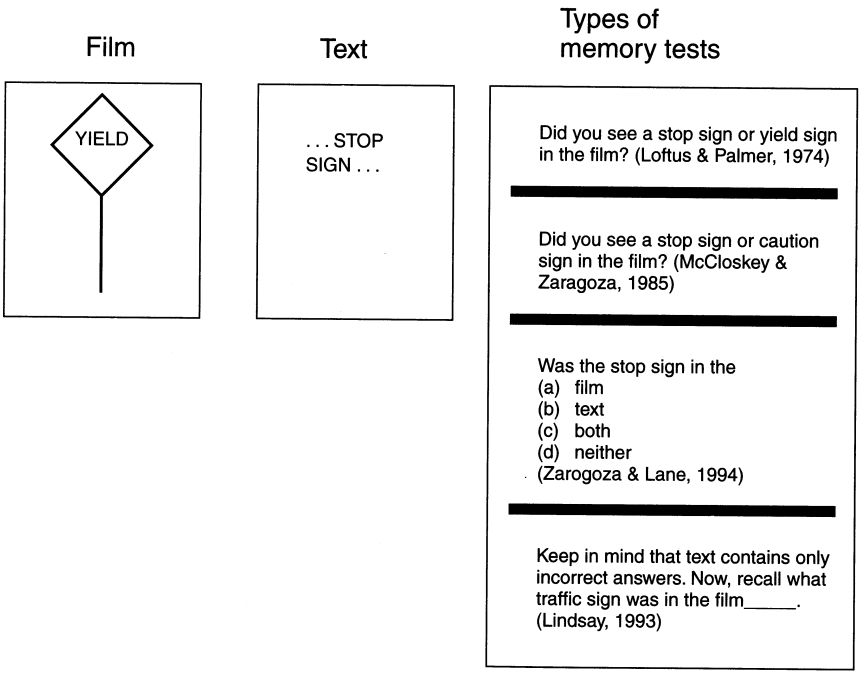


Figure 14.4
Experimental paradigms to investigate the impact of misleading information on eyewitness memory.

film, ask subjects to visualize the text, and present the memory test the next day. Source memory also tends to be poor if the misleading suggestions contained in the text are repeated several times rather than presented in the text only once (Zaragoza & Mitchell, 1996).

One other paradigm (see figure 14.4) that suggests that memory really is affected by subsequently presented misinformation comes from Lindsay (1990; also Weingardt, Loftus, & Lindsay, 1995; see Lindsay, 1993). Again, say that subjects see a yield sign in the film and later read a text about a stop sign. The description of a stop sign is misinformation. Now subjects are correctly informed that the text did not contain any correct answers to a subsequent memory test. In the memory test, subjects are asked to report details about the traffic sign, and subjects know that the correct answer comes from the film and not from the text. When it is hard to discriminate between the film and text experiences, subjects are likely to recall incorrectly details from the text, such as that the traffic sign was a stop sign. Such incorrect recall occurs less often when subjects are asked to recall details from the film about which no misinformation was given in the text. Here the demand characteristics of the experiment unambiguously push subjects to recall only the information in the film, yet they frequently recall inaccurately the information in the text.

I went through these experimental paradigms in some detail because I want you to see exactly how researchers refine their paradigms in response to alternative interpretations of findings. In this case, we can say that misinformation

is likely to affect recollection when it is relatively difficult for people to discriminate between the misinformation event and the to-be-remembered event (Lindsay, 1993). So, for example, our hypothetical Jim—mentioned at the beginning of this chapter—might be prone to remember incorrectly that the thief stole a camera (when, in fact, he saw the thief steal a radio) because he heard someone at the crime scene tell the police that the thief stole a camera. But Jim would be far less likely to remember incorrectly that the thief stole a camera if he heard someone talking about a missing camera the day after the theft and in a different physical setting than where the theft took place.

Loftus (1986) estimates that thousands of people in the United States are wrongfully convicted each year, and that many of these wrongful convictions are due to inaccurate eyewitness testimony. Juries deliberating the fate of people accused of crimes need to be made aware of the fallibility of human memory and the ease with which details of the past can be inaccurately recollected.

Hypnosis and Memory Sometimes it is supposed that hypnosis can help people better recollect crimes and accidents. As it turns out, psychologists debate whether hypnosis is a distinctive state of waking consciousness that is different from ordinary wakefulness or is merely an occasion in which some people are unusually motivated to carry out the requests of the hypnotist (see Farthing, 1992). Whatever the exact nature of hypnosis, certainly it is commonly believed that hypnosis promotes such accurate recall of the past that nearly all events must be stored in memory. The reality, though, is that when hypnosis is used to help eyewitnesses recollect a crime, accident, or any past event, hypnotized people do not remember details any more accurately than do nonhypnotized people. Hypnotized people, though, may be more confident about their recollections than nonhypnotized people (Buckhout, Eugenio, Licitia, Oliver, & Kramer, 1981; Krass, Kinoshita, & McConkey, 1989). Furthermore, hypnotized eyewitnesses are influenced by misleading questions even more than are nonhypnotized people.

Putnam (1979) presented subjects a videotape of a car accident and later hypnotized some subjects. When asked a misleading question like "Did you see the license plate number on the car?" some hypnotized subjects claimed to remember the numbers on the license plate when, in fact, the license plate was not visible in the film. Note that by using the phrase "the license plate," the question implies that the license plate was visible. Some of the hypnotized subjects presumably used the misleading implication in the question to reconstruct a number for the license plate. The subjects who were not hypnotized were less likely to fall for the misleading questions.

Hypnosis has also been used to attempt age regression, in which hypnotized adults may claim that they are really reliving some experience from childhood. But investigations reveal that the recollected details are often inaccurate (Nash, 1987). In one case, an adult who was hypnotically age-regressed remembered inaccurately a first-grade teacher's name. In another case, an adult who was hypnotically regressed to age 6 was asked to draw a picture. While the picture the adult produced looked childlike, it did not resemble the subject's own drawings made at age 6. Instead, the drawing reflected an adult's conception of a childish drawing, but not real children's drawings.

In brief, hypnosis, which is supposed to help people relive past experiences, does not really work. The research on hypnosis and memory is consistent with the idea that records of past experiences are not routinely maintained in memory, but must be reconstructed.

The Influence of Beliefs on Memory The idea that recollecting is reconstructing suggests that we reconstruct a memory of our past from our current beliefs and what we believe to be true about human personality in general (Ross, 1989). One idea that people have about personality is that beliefs remain rather stable over time. As a result, people tend to remember that their past beliefs were similar to their currently held beliefs, even when their beliefs have, in fact, changed over time. Let me provide a few experimental demonstrations.

In one study (Goethals & Reckman, 1973; see also Markus, 1986), high school students filled out a survey asking them for their opinion on various topics, including forced busing. About two weeks later, students met with a respected high school senior who presented a carefully crafted and well-rehearsed argument to the students about busing that was the opposite of the students' own opinion. For example, students who were opposed to forced busing heard a counterargument in favor of forced busing. Following the counterargument, students were again asked their opinion on busing, and were also asked to try to recall how they had filled out the survey two weeks earlier. The instructions emphasized the importance of accurate recall.

The counterarguments were effective; students tended to reverse their opinion about busing after hearing the counterargument. The result, consistent with reconstruction, was that the students tended to remember that they originally filled out the survey question about busing in a way consistent with their newly formed opinion and inconsistent with the way they actually had originally answered the busing question. For example, the students who were originally opposed to forced busing but heard a persuasive argument in favor of forced busing tended to remember that they had been in favor of forced busing all along. It was as if the students examined their current belief about busing, assumed that attitudes remain stable over the short period of two weeks, and so reconstructed that they must have held their current attitude two weeks earlier.

Galotti (1995) studied the criteria students use when selecting a college. Galotti asked students to recall the criteria that they had listed in a previously filled-out questionnaire assessing the basis on which they decided where to go to college. Galotti also asked the students to describe the ideal criteria that they thought, in retrospect, they ought to have used. The questionnaires had been filled out 8 to 20 months earlier. Subjects recalled about half of the criteria they had used when originally making the decision about where to go to college. But the overlap between what they recalled and the ideal criteria was substantially greater than the overlap between what they recalled and the criteria they had actually used. It was as if subjects used their current sense of the ideal decision criteria to reconstruct a memory of the criteria they used when originally making the college decision.

Many people, at least in our culture, believe that a woman's mood is likely to become more negative just before and during menstruation. It turns out,

though, that this belief may be false. Based on diary studies, there seems to be no reliable correlation between a woman's mood and her menstrual cycle, at least when large numbers of women are studied (see Ross, 1989). The idea of reconstruction suggests that women may use this belief about mood and menstruation to remember inaccurately that their mood had been worse during a previous menstruation phase than during an intermenstrual phase of the cycle.

Evidence consistent with the reconstruction hypothesis is provided by Ross (1989). He reports a study in which a group of women was asked to keep detailed diaries in which they recorded various life events and daily moods. The women were not told that the study focused on the menstrual cycle. At one point in the experiment, the women were asked to recall their mood from a day two weeks earlier. The women were supplied the date and day of the week and a small portion of their diary entries, including an entry that indicated whether they were menstruating. For one group, the to-be-recalled day was during the menstruation phase of their cycle, while for the other group the to-be-recalled day was during the intermenstrual phase. The actual diary entries for the to-be-recalled days indicated that the women's mood was no worse on average during the menstrual phase than during the intermenstrual phase. Yet the women tended to recall that their mood was worse on the menstrual day. Moreover, the more the women believed in a correlation between menstruation and mood (as assessed by an attitude survey), the more likely they were to exaggerate how negative their mood was on the day when they were menstruating.

Confidence and Accuracy

As I suggested earlier, record-keeping theories of human memory may concede that recollection of the past often involves reconstruction. The record-keeping theory could claim that a person resorts to reconstruction when the retrieval process fails to locate the necessary record. The constructionist theory claims instead that people use a reconstruction strategy every time they reflect on the past.

The record-keeping theory implies that people should be able to tell the difference between when they are able to read a record that accurately preserves the details of the past event, and when they are unable to locate the record and so must resort to making guesses about the past. People would presumably have more confidence in their memory for a past event if they are reading the record than if they are only reconstructing it. Therefore, according to the record-keeping theory, people's confidence in the accuracy of their memory for a past event should be reliably greater when the event is remembered accurately than when an event is remembered inaccurately.

The constructionist theory, on the other hand, claims that all recollection is reconstruction. Constructionist theory suggests that confidence and accuracy may sometimes be related, particularly when people have developed learning and reconstruction strategies for which they have been provided feedback as to how well those strategies work. In such cases, people may use their knowledge about how well a strategy has worked in the past to predict accurately how well it will work in the future. However, constructionist theory predicts that confidence will not be strongly related to accuracy when people have had no opportunity to develop adequate learning and reconstructive strategies, or

when there is misleading information that fools people into thinking that they have accurately reconstructed an event. In these latter two situations, people may be as confident in the accuracy of an incorrectly reconstructed event as they are of a correctly reconstructed event.

Consistent with the predictions of constructionist theory, a variety of experiments have demonstrated that the correlation between confidence and accuracy is typically quite low, especially in situations where eyewitnesses to crimes and accidents must recollect details of those crimes and accidents (Wells & Murray, 1984; Donders, Schooler, & Loftus, 1987; Smith, Kassin, & Ellsworth, 1989). Presumably most people have not had much practice developing learning and reconstruction strategies for eyewitness information, and therefore have not learned when such strategies produce accurate recollections (see Perfect, Watson, & Wagstaff, 1993).

On the other hand, the correlation between confidence and accuracy is reliably higher in situations where people have had such practice. For example, the correlation between confidence and accuracy is moderately high when subjects are asked to answer general knowledge questions, such as "Who wrote *The Mill on the Floss*?" (e.g., Hart, 1967; Perfect et al., 1993; see Nelson, 1988, for a review). Presumably most people have learned how good they are at answering general knowledge questions (Perfect et al., 1993). The correlation between confidence and accuracy is also moderately high when subjects are asked to answer questions about short texts they have recently read (e.g., Stephenson, 1984; Stephenson, Clark, & Wade, 1986). In this case, experience in academic settings has presumably taught most people how good they are at answering questions about texts.

Record-keeping theories of memory would predict that any variable that decreases memory accuracy should also decrease confidence in the accuracy of the memory. Contrary to the record-keeping prediction, Chandler (1994) reported a series of studies in which accuracy was decreased but confidence increased. Chandler had subjects study nature pictures, such as pictures of lakes. Later, subjects were required to determine which of two related pictures (e.g., two different lakes) had been previously displayed and to indicate their confidence in their recognition judgment. When the subjects had also studied a third related picture (e.g., a third lake), their recognition performance declined but their confidence in their selection increased (compared to the case when there was no third picture). The constructionist explanation is that subjects become more familiar with the general theme (e.g., scenic lakes) of the pictures as they study more of the related pictures. Both alternatives on the recognition test fit the theme, making discrimination between them difficult, so recognition memory performance declines. But because the selected picture fits the theme, confidence in the selection is high.

Also consistent with constructionist theory is the finding that people become confident of inaccurate recollections when those recollections are reconstructed from misleading information supplied to them by an experimenter (e.g., Davis & Schiffman, 1985; Spiro, 1977). Consider a study by Ryan and Geiselman (1991). They presented subjects a film of a robbery and a week later had them read a summary description of the film. For some of the subjects the summary included a misleading detail, such as "The police car is at a brown house" (in

fact, the house in the film was white). The subjects then answered questions about the film (e.g., "What was the color of the house?"). The interesting finding was that subjects who were biased by the incorrect detail, and therefore gave the wrong answer (e.g., "The house was brown"), were more confident of their wrong answer than were the subjects who were not given the misleading detail and so usually gave the correct answer (e.g., "The house was white").

People may become confident of their inaccurate memories when some inaccurately remembered piece of information is nevertheless consistent with the gist of some previously presented information. For instance, Roediger and McDermott (1995; see also Deese, 1959) presented subjects list of words (e.g., *bed, rest, awake*) for which every word on a list was related to a target word (e.g., *sleep*) not presented on the list. Later, on tests of recall and recognition, subjects remembered that the target words (e.g., *sleep*) were on the list about as often and with about the same confidence as they remembered the words that were actually presented on the lists. Moreover, the greater the number of related words presented on the list, the more likely subjects were to recall or recognize the target word not presented on the list (Robinson & Roediger, 1997). Presumably, the tendency to think of the target word when studying the list created a false memory for that target word that seemed as real to subjects as their memories of actually presented words.

The Overlap Principle

The fact that memory makes use of reconstruction strategies, such as relying on one's current beliefs to deduce past beliefs, means that remembering is often inaccurate. But recollections of the past are not inevitably inaccurate. The study of memory has established that memory of an event is more accurate when the environment at the time of recollection resembles the environment of the originally experienced event (Begg & White, 1985; Guthrie, 1959; Tulving, 1983; Tulving & Thomson, 1973). This principle may be called the *overlap principle*—people's memory for a past event improves to the extent that the elements of the recollection environment overlap with the elements of the past event. By environment, I mean a person's cognitive and emotional state, as well as the person's physical environment. The overlap principle also goes by the name of *encoding specificity*, to emphasize that how an event is processed or "encoded" will determine what kinds of cues will later be effective at promoting memory for the event (Tulving & Thomson, 1973).

Experimental Evidence for the Overlap Principle A good experimental demonstration of the overlap principle comes from research designed to help eyewitnesses more accurately remember crimes and accidents. Courts of law place strong emphasis on eyewitness accounts when assessing responsibility and punishment. Yet people often have a hard time remembering important details of crimes and accidents they have witnessed, a point I used earlier to illustrate the concept of reconstruction in memory. A variety of research suggests that eyewitness memory improves if the context surrounding the event is reinstated (see Geiselman, 1988).

Cutler and Penrod (1988; see also Geiselman, Fisher, MacKinnon, & Holland, 1985) had subjects view a videotape of a robbery and a few days later pick out

the robber from a lineup. Some subjects were given photographs (not containing the robber) taken from the scene of the crime, or were asked to think back through the events from beginning to end while imagining the robbery. These subjects tended to identify the robber more accurately than subjects not given any context-reinstating cues.

Other experiments have demonstrated that memory for an event is more accurate if retrieval takes place in the same physical environment as the one where the event originally occurred (e.g., Canas & Nelson, 1986; see Smith, 1988 for a review). In one of my favorite studies, subjects who learned a list of words while scuba diving later recalled more of the words if the recall test took place while the subjects were again scuba diving than if the recall test took place on land (Godden & Baddeley, 1975).

It should be noted, though, that the overlap of physical environments is probably an important determinant of memory when the to-be-learned information can be associated with the physical environment (Baddeley, 1982; Fernandez & Glenberg, 1985). An eyewitness may be more likely to remember the events of an accident, such as a car crashing into a tree, if the eyewitness recollects at the scene of the accident, than if the eyewitness recollects in the police station. The tree at the crash site is associated with the accident, so that seeing the tree is likely to activate information that may be used to reconstruct the accident. On the other hand, it is probably not as important that a student take an exam in the same room where he or she studied for the exam (Saufley, Otaka, & Bavaresco, 1985) since academic information would not ordinarily be associated with the physical elements of a room. Much more important is that the student understand the academic material, organize the material, and make the details contained within the material distinctive.

One demonstration that the overlap principle depends more on the similarity of cognitive processing than on similarity of physical stimuli comes from research on mood. The usual finding is that subjects induced to feel elated or depressed will more likely and quickly recall past events experienced in the same mood, than those experienced in a different mood (Snyder & White, 1982; Teasdale & Fogarty, 1979; see Blaney, 1986, for a review). In experiments conducted by Eich (1995), subjects were placed in a setting (e.g., a laboratory) and then responded to a list of 16 words designed to prompt memories of past experiences. The subjects' mood was also measured. Later, subjects were placed in either the same physical setting or a different setting, and were induced to feel either happy or sad. Mood was induced by having subjects listen to either joyful musical pieces while entertaining elating thoughts or melancholy musical pieces while entertaining depressing thoughts. The subjects then had to recall the 16 prompt words and the events the prompts elicited. Recall was better when the mood at the time of recall matched the mood experienced when the prompt words were first presented. Overlap in physical setting, on the other hand, did not matter to recall.

Problem Solving and the Overlap Principle Another interesting demonstration that the overlap principle is based on similarity in the way events are processed, and not on the mere presence of overlapping stimulus cues, comes from research on problem solving. A seemingly perplexing finding of this research is

that people often fail to remember facts that would help them solve a problem (Perfetto, Bransford, & Franks, 1983; Weisberg, Dicamillo, & Phillips, 1978). To illustrate with a hypothetical example, suppose a student in a psychology class learned the fact that, paradoxically, ignoring a young child who is whining and crying promotes the development of a dependent personality. On an examination, the student remembers this information and so correctly answers questions based on it. Yet when the student becomes a parent and encounters the whining of the child, the parent chooses to ignore the child, in the mistaken belief that the child will thereby become more independent.

Why does the parent fail to remember and make use of the relevant information previously learned in school? The answer is that trying to solve problems is unlikely to engage the portion of the cognitive system used to memorize facts, so the memorized facts play no role in the attempt to arrive at a solution. Perhaps if the parent had practiced solving child-rearing problems in school, rather than only memorizing facts about child rearing, the parent would have been more likely to transfer the information to real problems.

My hypothetical example about child rearing was inspired by experiments conducted by Adams et al. (1988); Perfetto et al. (1983); Lockhart, Lamon, and Glick (1988), and Needham and Begg (1991), among others. In one of these experiments (Perfetto et al., 1983), one group of subjects read a list of sentences that included sentences like "A minister marries several people a week" while a control group of subjects did not read the sentences. Both groups of subjects were then asked to solve brain teasers like "How can it be that a man can marry several women a week, never get divorced, yet break no law?" Note that the sentences the first group read were designed to help them solve the problems presented later on. Surprisingly, the subjects who first read the helpful sentences were no more likely to solve the brain teasers than the control subjects.

What would it take to get the subjects to make use of previously studied information to solve a new problem? Adams et al. (1988) and Lockhart et al. (1988) presented groups of subjects with sentences like this: "The man married ten people each week" and, 5 seconds after each sentence, gave the subjects a clue to help solve the puzzle suggested by the sentence—for example "a minister." Note that the subjects were not memorizing the sentences but instead approaching each sentence as a kind of miniature problem for which they were quickly given the solution. Subjects asked to approach the sentences as a set of problems were later on better at solving the brain teasers than either the subjects who first only memorized the sentences or the control subjects who never read any sentences. When the experiment ensured that the information processing activity required by the brain teaser matched the information processing activity required by the original presentation of the sentences, the subjects were able to make use of the sentences to help them solve the brain teasers.

These findings are examples of transfer appropriate processing (discussed earlier in the chapter), and are a manifestation of the overlap principle. If the kind of cognitive processing taking place in the testing environment resembles that taking place in the original learning environment, then what is learned will likely transfer to the test. Other experiments demonstrating transfer appropriate processing can be found in Blaxton (1989); in Glass, Krejci, and Goldman (1989); and in the Morris, Bransford, and Franks (1977) study I discussed in the

"levels of processing" section earlier in this chapter. The obvious educational implication of transfer appropriate processing is that if schools want to increase the odds that what students learn in school will help them solve problems later in life, then schools should engage students in solving problems that resemble those encountered outside of school. Students who demonstrate on examinations that they remember the material are not necessarily going to be able to use the material to solve problems they encounter later on.

Recognition versus Recall Another demonstration of the overlap principle is the finding that, under most circumstances, people can recognize more accurately than they can recall a past event (McDougall, 1904). A recognition test usually supplies more information about the original event than does a recall test, because the correct answer to any memory question is contained in the recognition test. For example, subjects asked to recall as many names as they could remember from their high school class that graduated 47 years earlier recalled on average only about 20 names (about 30% of the class), but accurately recognized about 45 names (about 65% of the class) (Bahrick, Bahrick, & Wittlinger, 1975).

It is possible, however, to devise situations for which people can recall what they are unable to recognize (Watkins & Tulving, 1975; see Klatzky, 1980, for a review). Such situations are characterized by a recall testing environment that more closely resembles the original learning environment than does the recognition testing environment.

In one experiment demonstrating recall without recognition (Nilsson, Law, & Tulving, 1988), subjects were presented a list of famous names (e.g., George Washington) in the context of descriptive phrases (e.g., "He was the first in a long line but the only one on horseback—George Washington"). Seven days later the subjects were given a recognition test in which a set of famous names was presented. This set contained the previously studied names as well as foils (e.g., Charles Darwin). Subjects had to indicate which names they had studied a week earlier. Then subjects were given the descriptive phrases and had to recall the famous names (e.g., "He was the first in a long line but the only one on horseback—?"). Subjects were often able to recall famous names that they did not recognize.

14.4 Forgetting

Forgetting past experiences, if not in their entirety, at least in most of their detail, seems the rule. Why do we so easily forget most of our past? Certainly, if we fail to pay attention to certain information contained in an event then we are unlikely to remember that information later on. Or if we are not motivated to try to remember an event, or are not given enough information to enable us to be sure what it is we are supposed to remember, then we are not likely to remember the event.

Interference

Another important reason for forgetting, besides those mentioned above, is that one's memory for any given event from one's past is undermined by the

occurrence of other events. When memory for an event is undermined by events that precede it, the result is called *proactive interference*. When memory for an event is undermined by events that follow it, the result is called *retroactive interference*.

A variety of experimental paradigms have been used to demonstrate interference (see Klatzky, 1980, or Watkins, 1979, for a review). The classic demonstration of retroactive interference comes from Jenkins and Dallenbach (1924). In their experiment, subjects were first presented a list of nonsense syllables, then spent the following 8 hours either asleep or awake, and then tried to recall the nonsense syllables. The subjects who remained awake, and so experienced more interfering events, recalled fewer syllables than did the subjects who went to sleep. This experiment is not the ideal demonstration of retroactive interference, though, because the subjects who experienced less interference also experienced a night of sleep. Maybe people are more motivated or less fatigued after sleeping, and so perform better on memory tests.

An example of a better controlled experiment demonstrating interference is provided by Kalbaugh and Walls (1973; also see Barnes & Underwood, 1959; McGeoch, 1942; Melton & Irwin, 1940). They required their eighth-grade subjects to study a critical passage describing the essential biographical details of a fictional character. Some subjects read no other passages while other subjects read either two or four other biographical passages. The additional passages were presented either before the critical passage or after the critical passage. As summarized in figure 14.5, the experiment demonstrated both retroactive and

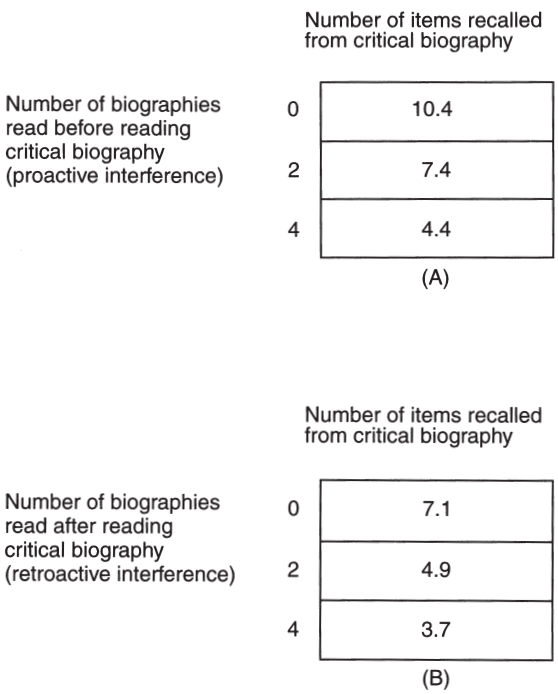


Figure 14.5
Proactive and retroactive interference. Based on Kalbaugh and Walls, 1973.

Number of facts about a location

1	2	3
1.169	1.196	1.248

Reaction time, in seconds, to respond "true"

Number of facts about a profession

1	2	3
1.144	1.202	1.267

Reaction time, in seconds, to respond "true"

Figure 14.6
The more facts about a professional or location, the longer the response time to verify the fact. Based on Anderson, 1974.

proactive interference. The subjects who read additional passages, whether presented before or after the critical passage, recalled less about the critical passage than subjects who read only the critical passage. The more additional passages the subjects read, the poorer was their recall of the critical passage.

Another way in which interference is demonstrated is in experiments that measure how fast people can decide whether a fact about a concept was on a previously memorized list of facts. Interference in this paradigm takes the form of an increased response latency to facts whose concepts are found in lots of other facts on the list. In one such experiment, subjects memorized sentences that described a professional in some location (e.g., "The lawyer was in the park"). The number of facts about any one professional or about any one location varied. Subjects might memorize two facts about a lawyer (e.g., "The lawyer was in the park," "The lawyer was at the beach") and one fact about a doctor (e.g., "The doctor was in the park") and might memorize two facts that involved parks and one fact that involved beaches. The typical result (illustrated in figure 14.6 for "true" responses) is that the more facts associated with a character or with a location, the longer it takes to decide if the fact is true or false (Anderson, 1974, 1976). Sometimes this finding is known as the *fan effect*. The more facts that "fan off" a concept, the longer it takes to verify whether any given fact about the concept was previously memorized.

Explaining Interference

The record-keeping theory has an easy explanation for interference. People search memory records by first finding in memory a target element, such as a character's profession. People then scan through the set of facts associated with the target element until the desired fact is found, or until the search is exhausted. The more associations to be searched, or the longer or more effort it takes to find the desired fact, the more likely the fact will not be found. It is as if

all of the associations, facts, and lists of facts stored in memory compete with the target information for the attention of the search process (Anderson, 1976, 1983; McGeoch, 1942; Postman, Stark, & Fraser, 1968).

The ease with which a record-keeping theory explains interference in memory experiments is one of the most compelling sources of evidence for it. But the explanation leads to a paradox (Smith, Adams, & Schoor, 1978). As we go through life, the number of associations with elements in our memory should continually increase. It follows, then, that over time we should become increasingly inefficient at finding information stored in our memory. Becoming an expert would be especially difficult, because the expert learns many facts about a set of concepts. Experts, then, would be expected to have an ever-increasing difficulty in remembering information in their area of expertise. Obviously this does not happen. The record-keeping theory's explanation for interference observed in many memorization experiments cannot easily explain the obvious facts that an adult's memory skill remains stable over time and that experts get better, and not worse, at remembering information in their area of expertise.

The difficulty that the record-keeping theory of memory has with explaining everyday observations about memory reminds us that explanations must have *ecological validity*. That is, theories of memory should explain how memory works in the actual environment in which we use our memory. Some memory theorists, notably Ulric Neisser (Neisser, 1978), have argued that a lot of memory research does not look at memory in realistic settings and so may not be ecologically valid. Neisser has urged the cognitive psychological community to make more use of experimental paradigms that resemble real-life situations. I have tried to include a fair number of such experiments in this chapter. Some memory theorists, however, have complained that experiments that have resembled real-life situations have not really uncovered any new principle of memory (Banaji & Crowder, 1989; see articles in the January 1991 issue of the *American Psychologist*). Perhaps, though, the contrast between the results of list memorization experiments and the everyday observations that memory is stable over time and that experts have good memory constitutes a compelling example of the importance of conducting ecologically valid research.

The constructionist theory is able to explain both the decline in memory performance exhibited in the memorization experiments and the lack of decline in memory observed in ordinary day-to-day situations or in experts. Memorization of related lists of words should generally present difficulties because the same elements would be used repeatedly to understand each new list. A subject who must memorize two lists of state-city associations, for example, would find that the connections among the cognitive elements used to understand and recollect the first list would be reconfigured when the second list was studied, thereby undermining memory for the details of the first list. Memorization, therefore, should be poorer or slower if a person has to memorize several related lists than if a person has to memorize unrelated lists. Furthermore, the repeated use of similar lists would make accurate and detailed reconstruction of any one list difficult.

The stability in an adult's memory skill occurs because the elements used to understand experiences do not expand in number as a result of having many experiences. Only the connections among elements change with experience.

Experts become good at remembering information in their area of expertise because the portions of their cognitive system that support the expertise will have many enduring patterns of connections to represent and reconstruct that information. For example, an expert on climate may be able to remember that Seattle has a milder winter than Denver by activating the general principle that oceans moderate climate. Much of what is involved in becoming an expert is understanding the general principles that a body of information entails.

One of the important predictions of the constructionist account of interference is that interference is not inevitable. Interference is expected when there is no effective learning strategy for extracting the patterns that integrate increasingly larger bodies of information. If such patterns can be extracted and used to reconstruct the information, then increasing the amount of information should not produce interference. Research confirms this prediction.

In one experiment, Jones and Anderson (1987) required subjects to memorize varying numbers of facts about hypothetical characters. In some cases, the facts were all related by a common theme. For example, subjects might learn that John has a rifle, John is a hunter, and John is in the forest. In other cases the facts were unrelated—for example: Jerry has a rifle, Jerry is a researcher, and Jerry is at the beach. As in other experiments investigating the fan effect, the subjects were then asked to verify whether particular facts were true (e.g., "John has a rifle") or false (e.g., "John is a researcher"). When the facts were unrelated, the usual fan effect was observed. Subjects took longer to verify facts about a character when there were six unrelated facts about that character to memorize than when there was only one fact to memorize. But when the facts were related by a common theme, the fan effect was greatly reduced. When the facts were related by a theme, subjects took about as long to verify facts about a character when there were six related facts to memorize about that character as when there was only one fact to memorize. Similar results have been obtained by Radvansky and Zacks (1991) and by Smith, Adams, and Schorr (1978). The constructionist explanation is that when the facts are related by a theme, that theme can be used to reconstruct whether a fact fits the theme and so must be true, or does not fit the theme, and so must be false.

The constructionist theory of memory also implies that interference depends on what kind of information subjects are asked to remember. If subjects in list-memorizing experiments were asked to remember the general pattern of information in the lists, rather than the unique details of each list, then presenting subjects with several lists should promote better memory of the general pattern, even while undermining memory for the unique details of each list.

Evidence consistent with the constructionist prediction is provided by Reder and Ross (1983). They required subjects to memorize a varying number of facts about hypothetical characters. Later, some subjects were required to indicate whether a particular fact was explicitly on the memorized list, while other subjects were asked to indicate whether a particular fact was similar to (implied by) other facts on the list. For example, subjects might learn three facts about Marvin: Marvin skied down the slope, Marvin waited in the lift line, and Marvin waxed his skis. The usual fan effect was observed if the memory test required subjects to judge whether a particular fact was explicitly on the list (e.g., Mar-

vin waited in the lift line). The greater the number of memorized facts about a character, the longer to verify whether any given fact about the character was on the list. The opposite of the usual fan effect was observed if the test required that subjects decide whether a fact was implied by other facts on the list (e.g., Marvin adjusted his skis). The greater the number of facts about a character, the faster subjects could verify whether a fact was similar to one of the memorized facts. The constructionist explanation is that the similarity judgment allowed subjects to compare a fact to the pattern or theme extracted from the memorized facts. The more facts there were to memorize, the more likely that such patterns would be extracted.

Another demonstration that increasing the amount of information can improve memory for patterns but undermine memory for details comes from an experiment by Bower (1974). Bower required his subjects to learn a critical passage describing the biography of a hypothetical character. The basic form of the biography described the time and place of birth, the occupation of the character's father, the way the father died, and so on. Subjects then studied some additional passages. For some subjects, the additional passages were biographies similar in form but different in detail from the critical passage. For other subjects, the additional passages were unrelated to the critical passage in form and in detail. Later, all subjects had to recall the same critical passage. Bower found that subjects who studied the related passages recalled fewer details of the critical passage (e.g., the father was a servant) but more of the general pattern of the passage (e.g., the passage described the father's occupation) than did subjects who studied unrelated passages.

Students often feel overwhelmed by the amount of material they must learn for an exam. Perhaps it would hearten them to learn that interference for newly acquired information is not inevitable. If students can relate each piece of information to a common theme, then interference is not likely to occur. The student should be able to remember large sets of information as well as small sets. Similarly, if the examination tests for general principles rather than specific details, then again interference is not inevitable. The more information one must learn, the more likely the general principles can be extracted from the information.

Summary and Conclusions

In the first section of this chapter, I introduced two competing types of theories of memory. One is the record-keeping theory, which argues that memory is a system for storing records of past events, that recollection is searching through and reading the records, and that forgetting is caused primarily by the distracting presence of many memory records. The second is the constructionist theory, which argues that memory reflects changes to the cognitive systems used to interpret events, that recollection is reconstructing the past, and that forgetting is caused primarily by the continuous changes each new experience makes to the cognitive systems that interpret and act on stimuli. Few contemporary theories of memory embody all the features of record-keeping theories, although some contemporary theories, especially those that use computers as metaphors for memory, seem closer in spirit to the record-keeping than to the

constructionist theory. Certainly the record-keeping theory has dominated the history of memory research and seems to reflect the ordinary person's view of memory (Loftus & Loftus, 1980).

I have argued that the evidence overall supports the constructionist theory over the record-keeping theory. In the second section, I discussed how experiences are retained in memory. Evidence consistent with constructionist theory is that memory is good for invariants or patterns that endure across many experiences, but is poor for the details of specific experiences. Usually people remember the details of a particular experience because those details are unusual or distinctive in some way. Even people with very remarkable memory for details, such as Luria's *S*, make use of mnemonic devices and learning strategies that help them make information more distinctive.

The constructionist approach claims that memory reflects the strength of connections among elements of the cognitive systems used to perceive, think about, and act on events. Such connections undergo continuous reconfiguration in response to experiences. In a sense, memory is only a byproduct of connections among the components of various cognitive systems. There is no separate memory system in which information is "stored." Consistent with the idea of memory as a byproduct is the assimilation principle: How well people remember new information about a topic depends on how much they already know about that topic. Also consistent is the observation that individual differences in memory are largely attributable to expertise in the relevant domain of knowledge. General intellectual skill, or skill at memorizing, does not seem to predict memory for new information from some domain of knowledge as well as does expertise in that domain.

Especially telling for the constructionist theory is that conscious recollection of the past depends on current knowledge and on recollection strategies. As I discussed in the third section, a person's recollections of the past are often distorted by misleading questions or general knowledge. For example, eyewitnesses to crimes and accidents sometimes mistakenly remember details, like a car going through a stop sign, that they never observed. Usually such mistakes are made when someone or some process implies that the details were a part of the crime or accident. Especially difficult for the record-keeping theory is the finding that people are often as confident of inaccurate as of accurate reconstructions of past events.

Although forgetting is common, people certainly are able to reconstruct accurately some of their past experiences. Memory is more accurate when there is considerable similarity between the retrieval and original learning environment, a phenomenon called the overlap (also called the encoding specificity) principle. The constructionist theory explains the overlap principle by claiming that memory is improved when the retrieval environment activates the same portions of the cognitive system used to interpret the original environment. For instance, people are more likely to use information previously learned in one environment to solve a new problem if the original environment also required them to use that information to solve problems.

The most important principle of forgetting, called interference, is that the more information a person must memorize, the more likely the person will be unable to remember or will be slower at remembering any given piece of

information. As I discussed in the fourth section, the record-keeping theory suggests that interference is primarily due to the distracting effects of other memory records, which increase in number as the amount of information to be remembered increases. But the record-keeping theory implies a paradox: Adults should show a gradual decline in their memory as they learn more about various topics. Experts should have especially poor memory in their domains of expertise. Yet neither of these propositions is true.

The constructionist theory predicts interference when no distinctive patterns enable the person to reconstruct information, as is likely to happen in list memorization experiments. Because the constructionist theory claims that no memory records are kept, an adult's memory remains stable over time. Because experts learn to find patterns in and to develop reconstruction strategies for their domain of expertise, experts have a good memory for that domain. The constructionist theory correctly predicts that interference usually observed in list learning experiments is eliminated if the memorized facts can be integrated by a common theme, or if the memory test requires people to remember the patterns rather than the details contained within the memorized material.

Although constructionist accounts of memory are currently influential (see Schacter, 1996), some cognitive psychologists continue to support record-keeping theories (see Hall, 1990). One might argue that, with suitable modifications, the record-keeping theory can explain the data I claimed support the constructionist theory. For example, a record-keeping theory could include a pattern recognition system that either stores descriptions of patterns or examines memory records to find patterns in events. Consequently, patterns of experiences would be readily remembered. A record-keeping theory could posit that reconstruction strategies are used when a sought-after memory record is not located.

It is true that such modifications would make the record-keeping theory work more like real human memory. Note, though, that the proposed modifications have the effect of making the record-keeping theory more like the constructionist theory. Furthermore, the modifications are not intrinsic to, or a natural consequence of, the central idea that memory is a matter of storing records of experiences. There is nothing about putting a record of an experience someplace in a storage bin that inevitably leads to extracting a pattern. There is nothing about reading memory records that leads to making plausible guesses about what happened in the past. These modifications are just tacked on, because without them the system does not resemble human memory. To put it another way, the record-keeping theory so modified lacks theoretical elegance.

In contrast, consider that the central idea of the constructionist theory, that the cognitive systems change the strength of their connections in response to events, does lead naturally to how human memory actually works. Remembering patterns, but not details, is a natural consequence of such a system, because the invariants in experiences strengthen already existing connections. No pattern recognition system has to be added on. Reconstruction happens because no records of past experiences are ever "read" or "reexperienced"; rather, past events must be inferred from the current state of connections. And a constructionist theory of memory more closely reflects what is known about the neurophysiology of learning and remembering.

Recommended Readings

Schacter's (1996) *Searching for Memory* is an outstanding book in which the author skillfully weaves theory, experimentation, real-life issues, and contemporary art in an exciting discussion of the current state of memory research. Neisser's historically important (1967) *Cognitive Psychology* includes a chapter on why memory is reconstructive and not reproductive; and his (1981) article in the journal *Cognition* discusses the theoretical implications of John Dean's memory of the Watergate coverup. Raaijmakers and Shiffrin (1992) provide a rigorous discussion of several theories of memory, including theories I label record keeping. Almost any study by Loftus, an enthusiastic advocate of constructionist approaches to memory, is informative and entertaining—try Loftus, Miller, and Burns (1978); Loftus (1979); or Weingardt, Loftus, and Lindsay (1995). Ross (1989) discusses several memory experiments, including the menstruation–mood experiment, in a review article assessing the implications of constructed memory for social attitudes and behaviors. J. Anderson's (1974, 1976) fan effect experiments remain elegant approaches to the study of memory by a talented scientist who happens to favor the record-keeping perspective.

References

Adams, J. L., Kasserman, J. E., Yearwood, A. A., Perfetto, G. A., Bransford, J. D., & Franks, J. J. (1988). Memory access: The effects of fact-oriented versus problem-oriented acquisition. *Memory and Cognition, 16*, 167–175.

Anderson, J. R. (1974). Retrieval of prepositional information from long-tern memory. *Cognitive Psychology, 5*, 451–474.

Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, D.C.: Winston.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*, 703–719.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2). New York: Academic Press.

Baddeley, A. D. (1982). Domains of recollection. *Psychological Review, 22*, 88–104.

Bahrick, H. P., Bahrick, P. C., & Wittlinger, R. P. (1975). Fifty years of memories for names and faces: A cross-sectional approach. *Journal of Experimental Psychology, 104*, 54–75.

Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist, 44*, 1185–1193.

Barclay, C. R., & Wellman, H. M. (1986). Accuracies and inaccuracies in autobiographical memories. *Journal of Memory and Language, 25*, 93–106.

Barnes, J. M., & Underwood, B. J. (1959). "Fate: of first-list associations in transfer theory. *Journal of Experimental Psychology, 58*, 97–105.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.

Beach, K. D. (1988). The role of external mneumonic symbols in acquiring an occupation. In M. M. Gruensberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1, pp. 342–346). Chichester, England: Wiley.

Begg, I., & White, P. (1985). Encoding, specificity in interpersonal communication. *Canadian Journal of Psychology, 39*, 70–87.

Bellezza, F. S., & Buck, D. K. (1988). Expert knowledge as mneumonic cues. *Applied Cognitive Psychology, 2*, 147–162.

Bjork, R. A. (1975). Short-term storage: The ordered output of a central processor. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindeman, & D. B. Pisoni (Eds.), *Cognitive Theory* (Vol. 1). Hillsdale, NJ: Erlbaum.

Blaney, P. H. (1986). Affect and memory: A review. *Child Development, 53*, 799–810.

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 657–668.

Bobrow, S., & Bower, G. H. (1969). Comprehension and recall of sentences. *Journal of Experimental Psychology, 80*, 455–461.

Bower, G. H. (1974). Selective facilitation and interference in retention of prose. *Journal of Educational Psychology, 66*, 1–8.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorical word lists. *Journal of Verbal Learning and Verbal Behavior, 8*, 303–343.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology, 2*, 331–350.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisite for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717–726.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3*, 193–209.

Bransford, J. D., McCarrell, N. S., Franks, J. J., & Nitsch, K. E. (1977). Toward unexplaining memory. In R. S. Shaw & J. D. Bransford (Eds.), *Perceiving, Acting and Knowing: Toward an Ecological Psychology*. Hillsdale, NJ: Erlbaum.

Brewer, W. F. (1984). The nature and function of schema. In J. Strachey & T. K. Srult (Eds.), *Handbook of Social Cognition* (Vol. 1, pp. 119–160). Hillsdale, NJ: Erlbaum.

Brooks, L. W., & Dansereau, D. F. (1983). Effects of structural schema training and text organization on expository prose processing. *Journal of Educational Psychology, 75*, 811–820.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*, 73–99.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomemon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325–337.

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology, 62*, 311–318.

Buckhout, R., Eugenio, P., Licitia, T., Oliver, L., & Kramer, T. H. (1981). Memory, hypnosis, and evidence: Research or eyewitnesses. *Social Action and the Law, 7*, 67–72.

Canas, J. J., & Nelson, D. C. (1986). Recognition and environmental context: The effects of testing by phone. *Bulletin of the Psychonomic Society, 24*, 407–109.

Carlson, N. R. (1994). *Physiology of behavior*. Boston: Allyn & Bacon.

Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory and Cognition, 22*, 273–280.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–81.

Chiesi, H., Spilich, G., & Voss, J. F. (1979). Acquisition of domain related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*, 257–273.

Cohen, M. E., & Carr, W. J. (1975). Facial recognition and the VonRestorff effect. *Bulletin of the Psychonomic Society, 6*, 383–384.

Collins, A. F., & Hay, D. C. (1994). Connectionism and memory. In P. E. Morrise & M. Gruneberg (Eds.), *Theoretical aspects of memory* (pp. 196–237). New York: Routledge.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268–294.

Dale, H. C. A., & Baddeley, A. D. (1962). On the nature of alternatives used in testing recognition of memory. *Nature, 196*, 93–94.

Davis, J., & Schiffman, H. R. (1985). The influence of the wording of interrogatives on the accuracy of eyewitness recollection. *Bulletin of the Psychonomic Society, 23*, 394–396.

Dean, R. S., & Kulhavy, R. W. (1981). Influence of spatial organization in prose learning. *Journal of Educational Psychology, 73*, 64–97.

DeAngelis, T. (1988). Dietary recall is poor: Recall study suggests. *APA Monitor, 19*, 14.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22.

DeMarie-Dreblow, D. (1991). Relation between knowledge and memory: A reminder that correlation does not imply causation. *Child Development, 62*, 484–498.

Donders, K., Schooler, J. W., & Loftus, E. F. (1987, November). *Troubles with memory*. Paper presented at the annual meeting of the Psychonomic Society, Seattle, WA.

Dooling, D. J., & Christiaansen, R. E. (1977). Episodic and semantic aspects of memory for prose. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 428–436.

Ebbinghaus, H. (1885). *Uber das Gedachtnis*. Leipzig: Dunker and Humbolt.

Ericssson, K. A., & Polson, P. G. (1988). An experimental analysis of the mechanisms of a memory skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 305–316.

Erickson, J. R., & Jemison, C. R. (1991). Relations among measures of autobiographical memory. *Bulletin of the Psychonomic Society, 29*, 233–236.

Farthing, G. W. (1992). *The psychology of consciousness*. Upper Saddle River, NJ: Prentice-Hall.

Fernandez, A., & Glenberg, A. M. (1985). Changing environmental context does not reliably affect memory. *Memory and Cognition, 13*, 333–336.

Freeman, L. C., Romney, A. K., & Freeman, S. C. (1987). Cognitive structure and informant accuracy. *American Anthropologist, 89*, 310–325.

Fries, E., Green, P., & Bowen, D. J. (1995). What did I eat yesterday? Determinants of accuracy in 24-hour food memories. *Applied Cognitive Psychology, 9*, 143–155.

Galiotti, K. M. (1995). Memories of a "decision-map": Recall of real-life decision. *Applied Cognitive Psychology, 9*, 307–319.

Geiselman, R. E. (1988). Improving eyewitness memory through mental reinstatement of context. In G. M. Davies & D. M. Thompson (Eds.), *Memory in context: Context in memory* (pp. 231–244). Chichester, England: Wiley.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mneumonics versus hypnonsis. *Journal of Applied Psychology, 70*, 401–412.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology, 17*, 324–363.

Glass, A. L., Krejci, J., & Goldman, J. (1989). The necessary and sufficient conditions for motor learning, recognition, and recall. *Journal of Memory and Language, 28*, 189–199.

Godden, N. N., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66*, 325–332.

Goethals, G. R., & Reckman, R. F. (1973). The perception of consistency in attitudes. *Journal of Experimental Psychology, 9*, 491–501.

Goldstein, A. G., & Chance, J. (1970). Visual recognition memory for complex configurations. *Perception and Psychophysics, 9*, 237–241.

Greeno, J. G. (1964). Paired-associate learning with massed and distributed repetition of items. *Journal of Experimental Psychology, 67*, 286–295.

Groninger, L. D. (1971). Mnemonic imagery and forgetting. *Psychonomic Science, 23*, 161–163.

Grossberg, S., & Stone, G. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review, 93*, 46–74.

Guthrie, E. R. (1959). Association of contiguity. In S. Koch (Ed.), *Psychology: A study of science* (Vol. 2). New York: McGraw-Hill.

Haber, R. N., & Haber, L. R. (1988). The characteristic of eidetic imagery. In L. K. Obler & D. Fein (Eds.), *The exceptional brain: Neuropsychology of talent and special abilities* (pp. 218–241). New York: Guilford Press.

Hall, J. F. (1990). Reconstructive and reproductive models of memory. *Bulletins of the Psychonomic Society, 28*, 191–194.

Hanawalt, N. G., & Demarest, I. H. (1939). The effect of verbal suggestion in the recall period upon the production of visually perceived forms. *Journal of Experimental Psychology, 25I*, 151–174.

Harris, R. O., & Monaco, G. E. (1978). Psychology of pragmatic implication: Information processing between the lines. *Journal of Experimental Psychology: General, 107*, 1–22.

Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior, 76*, 685–691.

Hoffman, R. R., Bringmann, W., Bamberg, M., & Klein, R. (1986). Some historical observations on Ebbinghaus. In D. Gorgein & R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus centennial conference*. Hillsdale, NJ: Erlbaum.

Holmes, D. S. (1972). Repression or interference: A further investigation. *Journal of Personality and Social Psychology, 22I*, 163–170.

Howes, M. B. (1990). *The psychology of human cognition: Mainstream and Genevan tradtionsI*. New York: Pergamon Press.

Hunt, E., & Love, T. (1972). How good can memory be? In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, DC: Winston.

Hyde, J. S., & Jenkins, J. J. (1975). Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior, 12*, 471–480.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17*, 649–667.

Jenkins, J. G., & Dallenbach, K. M. (1924). Oblivescence during sleep and working. *American Journal of Psychology, 35*, 605–612.

Jones, W. P., & Anderson, J. R. (1987). Short- and long-term memory retrieval: A comparison of the effects of information load and relatedness. *Journal of Experimental Psychology: General, 116*, 137–153.

Kalbaugh, G. L., & Walls, R. T. (1973). Retroactive and proactive interference in prose learning of biographical and science materials. *Journal of Educational Psychology, 65*, 244–251.

Klatzky, R. L. (1980). *Human memory: Structure and processes*. San Francisco: Freeman.

Koriat, A., & Melkman, R. (1987). Depth of processing and memory organization. *Psychological Records, 49*, 183–187.

Krass, J., Kinoshita, S., & McConkey, K. M. (1989). Hypnotic memory and confidence reporting. *Applied Cognitive Psychology, 3*, 35–51.

Kuhara-Kojima, K., & Hatano, G. (1991). Contribution of content knowledge and learning ability to the learning of facts. *Jornal of Educational Psychology, 83*, 253–263.

Lindsay, D. S. (1990). Misleading questions can impair eyewitnesses' ability to remember details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1077–1083.

Lindsay, D. S. (1993). Eyewitness suggestibility. *Current Directions in Psychological Science, 3*, 86–89.

Linton, M. (1978). Real world memory after six years: An in vivo study of very long term memory. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory*. Orlando, FL/London: Academic Press.

Lockhart, R. S., Lamon, M., & Glick, M. L. (1988). Conceptual transfer in simple insight-problems. *Memory and Cognition, 16*, 36–44.

Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.

Loftus, E. F. (1980). *Memory*. Menlo Park, CA: Addison-Wesley.

Loftus, E. F. (1982). Remembering recent experiences. In L. S. Cermak (Ed.), *Human memory and amnesia*. Hillsdale, NJ: Erlbaum.

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist, 48*, 518–537.

Loftus, G. R., & Loftus, E. F. (1980). The influence of one memory retrieval on a subsequent memory retrieval. *Memory and Cognition, 2*, 467–471.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Memory, 13*, 585–589.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 19–31.

Lorayne, H., & Lucas, J. (1974). *The memory book*. New York: Ballantine.

Lorch, R. F., & Lorch, E. P. (1985). Topic structure representation and text recall. *Journal of Educational Psychology, 77*, 137–148.

Luria, A. R. (1968). *The mind of mnemonist*. New York: Basic Books.

Maki, R. (1989). Recognition of added and deleted details in scripts. *Memory and Cognition, 17*, 274–282.

Mandler, J. M. (1978). A code in the node: The use of a story schema in retrieval. *Discourse Processes, 1*, 14–35.

Markus, G. B. (1986). Stability and change in political attitudes: Observe, recall, and "explain." *Political Behavior, 8*, 21–44.

Mayer, R. E. (1980). Elaboration techniques that increase the meaningfulness of technical text: An experimental test of the learning strategy hypothesis. *Journal of Educational Psychology, 72*, 770–784.

Mayer, R. E., & Bromage, B. D. (1980). Different recall protocols for technical tests due to advance organizers. *Journal of Educational Psychology, 72*, 209–225.

McCloskey, M., & Zaragoza, M. (1985). Misleading posteevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General, 114*, 1–16.

McCloskey, M., Wible, C. G., & Cohen, N. J. (1988). Is there a special flash-bulb memory mechanism? *Journal of Experimental Psychology: General, 117*, 171–181.

McDougall, R. (1904). Recognition and recall. *Journal of Philosophical and Scientific Methods, 1*, 229–233.

McGeoch, J. A. (1942). *The psychology of human learning*. New York: Longmans, Green.

Melton, A. W., & Irwin, J. M. (1940). The influence of degrees of interpolated learning on retroactive inhibitions and the overt transfer of specific responses. *Journal of Experimental Psychology, 53*, 173–203.

Morris, P. (1998). Memory research: Past mistakes and future prospects. In G. Claxton (Ed.), *Growth points in cognition*. London: Routledge.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519–534.

Nash, M. (1987). What, if anything, is regressed about hypnotic age regression? A review of the literature. *Psychological Bulletin, 102*, 42–52.

Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory and Cognition, 19*, 543–557.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Neisser, U. (1978). Memory: What are the important questions? In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 3–24). London: Academic Press.

Neisser, U. (1981). John Dean's memory: A case study. *Cognition, 9*, 1–22.

Neisser, U., & Harsch, N. (1991). Phantom flashbulbs: False recognition of hearing the news about the *Challenger*. In E. Winnograd & U. Neisser (Eds.), *Flashbulb memories: Recalling the Challenger explosion and other disasters*. New York: Cambridge University Press.

Nelson, T. O. (1988). Predictive accuracy of feeling of knowing across different criterion tasks and across different subject populations and individuals. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1, pp. 190–196). Chichester, England: Wiley.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.

Parkin, A. J. (1984). Levels of processing, context, and facilitation of pronunciation. *Acta pscyhologica, 55*, 19–29.

Parry, H., & Crossley, H. (1950). Validity of responses to survey questions. *Public Opinion Quarterly, 14*, 61–80.

Penfield, W. W. (1969). Consciousness, memory, and man's conditioned reflexes. In K. Pribram (Ed.), *On the biology of learning*. New York: Harcourt, Brace and World.

Penfield, W. W., & Jasper, H. (1954). *Epilepsy and the functional anatomy of the human brain*. Boston: Little, Brown.

Penfield, W. W., & Perot, P. (1963). The brain's record of auditory and visual experience. *Brain, 86*, 595–696.

Perfect, T. J., Watson, E. L., & Wagstaff, G. F. (1993). Accuracy of confidence ratings associated with general knowledge and eyewitness memory. *Journal of Applied Psychology, 78*, 144–147.

Perfetto, G. A., Bransford, J. D., & Franks, J. J. (1983). Constraints on access in a problem-solving context. *Memory and Cognition, 11*, 24–31.

Pezdek, K., Maki, R., Valencea-Lover, D., Whetstone, T., Stoeckert, J., & Dougherty, T. (1988). Picture memory: Recognizing added and deleted details. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 468–476.

Pillemer, D. B. (1984). Flashbulb memories of the assassination attempt on President Reagan. *Cognition, 16*, 63–80.

Posner, M. I., & Warren, R. E. (1972). Traces, concepts, and conscious constructions. In A. W. Melton & T. E. Martin (Eds.), *Coding processes in human memory*. Washington, DC: Winston.

Postman, L., Stark, K., & Fraser, J. (1968). Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior, 7*, 672–694.

Pressley, M., & Brewster, M. E. (1990). Cognitive elaboration of illustrations to facilitate acquisition of facts: Memories Prince Edward School. *Applied Cognitive Psychology, 4*, 359–369.

Pressley, M., & Van Meter, P. (1994). What is memory development the development of? In P. E. Morris & M. Cruneberg (Eds.), *Theoretical aspects of memory* (pp. 79–129). London: Routledge.

Putnam, B. (1979). Hypnosis and distortions in eyewitness memory. *International Journal of Clinical and Experimental Hypnosis, 27*, 437–448.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1992). Models for recall and recognition. *Annual Review of Psychology, 43*, 205–234.

Radavansky, G. A., & Zacks, R. T. (1991). Mental models and the fan effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 940–953.

Reder, L. M., & Ross, B. H. (1983). Integrated knowledge in different tasks: The role of retrieval strategy on fan effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 55–72.

Robbins, L. C. (1963). The accuracy of parental recall of aspects of child development and of child rearing practices. *Journal of Abnormal and Social Psychology, 66*, 261–270.

Robinson, K. J., & Roediger, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8*, 231–237.

Roediger, H. L. (1980). Memory metaphors in cognitive psychology. *Memory and Cognition, 8*, 231–246.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814.

Rosenfield, I. (1988). *The invention of memory*. New York: Basic Books.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review, 96*, 341–357.

Ross, D. F., Ceci, S. J., Dunning, D., & Toglia, M. P. (1994). Unconscious transference and mistaken identity: When a witness misidentifies a familiar but innocent person. *Journal of Applied Psychology, 79*, 990–892.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representation by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 216–271). Cambridge, MA: MIT Press.

Ryan, R. H., & Geiselman, R. E. (1991). Effects of biased information on the relationship between eyewitness confidence and accuracy. *Bulletin of the Psychonomic Society, 29*, 7–9.

Sachs, J. D. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics, 2*, 437–442.

Saufley, W. H., Otaka, S. R., & Bavaresco, J. L. (1985). Context effects: Classroom tests and context independence. *Memory and Cognition, 13*, 522–528.

Schacter, D. L. (1996). *Searching for memory*. New York: Basic Books.

Schmidt, S. R. (1985). Encoding and retrieval processes in the memory for conceptually distinctive events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 565–578.

Schmidt, S. R. (1991). Can we have a distinctive theory of memeory? *Memory and Cognition, 19*, 523–542.

Schneider, W., Korkel, J., & Weinert, F. E. (1987). *The knowledge base and memory performance: A comparison of academically successful and unsuccessful learners*. Paper presented at the meeting of the American Educational Research Association, Washington, DC.

Searleman, A., & Hermann, D. (1994). *Memory from a broader perspective*. New York: McGraw-Hill.

Shephard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*, 156–163.

Smith, E. E. (1988). Concepts and thought. In R. J. S. Sternberg & E. E. Smith (Eds.), *The psychology of human thought*. Cambridge, England: Cambridge University Press.

Smith, A. D., & Winograd, E. (1978). Adult age differences in remembering faces. *Developmental Psychology, 14*, 443–444.

Smith, E. E., Adams, N., & Schorr, D. (1978). Fact retrieval and the paradox of interference. *Cognitive Psychology, 10*, 438–464.

Smith, V. L., Kassin, S. M., & Ellsworth, P. C. (1989). Eyewitness accuracy and confidence: Within versus between-subjects correlations. *Journal of Applied Psychology, 74*, 356–359.

Smith, A. F., Jobe, J. B., & Mingay, D. J. (1991). Retrieval from memory of dietary information. *Applied Cognitive Psychology, 5*, 269–296.

Snyder, M., & Uranowitz, S. W. (1978). Reconstructing the past: Some cognitive consequences of person perception. *Journal of Personality and Social Psychology, 36*, 941–950.

Snyder, M., & White, P. (1982). Moods and memories: Elation, depression, and the remembering of the events of one's life. *Journal of Personality, 50*, 142–167.

Spiro, R. J. (1977). Remembering information from text: The state of the schema approach. In R. C. Anderson, R. J. Spiro, & W. E. Monague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Erlbaum.

Sporer, S. L. (1991). Deep-deeper-deepest? Encoding strategies and the recognition of human faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 323–333.

Squire, L. R. (1987). *Memory and brain*. New York: Oxford University Press.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology, 25*, 207–222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single trial learning of 2560 visual stimuli. *Psychonomic Science, 19*, 73–74.

Stein, B. S., & Bransford, J. D. (1979). Constraints on effective elaboration: Effects of precision and subject generation. *Journal of Verbal Learning and Verbal Behavior, 18*, 769–777.

Stein, B. S., Littlefield, J., Bransford, J .D., & Persampieri, M. (1984). Elaboration and knowledge acquisition. *Memory and Cognition, 12*, 522–529.

Stephenson, G. M. (1984). Accuracy and confidence in testimony: A critical review and some fresh evidence. In D. J. Muller, D. E. Blackman, & A. J. Chapman (Eds.), *Psychology and law: Topics from an international conference* (pp. 229–249). Chichester, England: Wiley.

Stephenson, G. M., Clark, N. K., & Wade, G. S. (1986). Meetings make evidence: An experimental study of collaborative and individual recall of a simulated police interrogation. *Journal of Personality and Social Psychology, 50*, 1113–1122.

Stromeyer, C. F. Ill, & Psotka, J. (1970). The detailed textures of eidetic images. *Nature, 225*, 346–349.

Sulin, R. A., & Dooling, D. J. (1974). Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology, 103*, 255–262.

Teasdale, J. D., & Fogarty, S. J. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology, 188*, 248–257.

Thompson, C. P., Cowan, T., Frieman, J., Mahadevan, R. S., & Vogel, R. J. (1991). Rahan: A study of a memorist. *Journal of Memory and Language, 30*, 702–724.

Thorndyke, P. W. (1976). The role of inferences in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior, 15*, 437–446.

Thorndyke, P., & Hayes-Roth, B. (1979) The use of schemata in the acquisition and transfer of knowledge. *Cognitive Psychology, 11*, 82–106.

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352–373.

Von Restorff, H. (1933). Uber die Wirkung von Bereichsbildungen in Spurenfeld. *Psychologisch Forschung, 18*, 299–342.

Walker, C. H. (1987). Relative importance of domain knowledge and overall aptitude on acquisition of domain-related information. *Cognition and Instruction, 4*, 24–42.

Waters, R., & Leeper, R. (1936). The relation of affective tone to the retention of experience in everyday life. *Journal of Experimental Psychology, 19*, 203–215.

Watkins, M. J. (1979). Engrams as cuegrams and forgetting as cue overload: A cueing approach to the structure of memory. In C. R. Puff (Ed.), *Memory organization and structure*. New York: Academic Press.

Watkins, M. J., & Tulving, E. (1975). When recognition fails. *Journal of Experimental Psychology: General, 104*, 5–29.

Webber, E. U., & Marshall, P. H. (1978). Priming in a distributed memory system. Implications for models of implicit memory. In S. Lewandowsky, J. C. Dunn, & K. Kirsner (Eds.), *Implicit memory: Theoretical Issues* (pp. 87–98). Hillsdale, NJ: Erlbaum.

Weingardt, K. R., Loftus, E. F., & Lindsay, D. S. (1995). Misinformation revisitedJ: New evidence on the suggestibility of memory. *Memory and Cognition, 23*, 72–82.

Weisberg, R., Dicamillo, M., & Phillips, D. (1978). Transferring old associations to anew problems: A nonautomatic process. *Journal of Verbal Learning and Verbal Behavior, 17*, 219–228.

Wells, G, L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological Perspectives* (pp. 155–170). New York: Cambridge University Press.

White, R. T. (1982). Memory for personal events. *Human Learning, 1*, 171–183.

White, R. T. (1989). Recall of autobiographical events. *Applied Cognitive Psychology, 3*, 127–135.

Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology, 9*, 418–455.

Winnograd, T. (1976). Computer memories: A metaphor for memory organization. In C. N. Cofer (Ed.), *The structure of human memory*. San Francisco: Freeman.

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 934–945.

Zaragoza, M. S., & Mitchell, K. J. (1996). Repeated exposure to suggestion and the creation of false memories. *Psychological Science, 7*, 294–300.

PART VIII

Attention

Chapter 15

Attention and Performance Limitations

Michael W. Eysenck and Mark T. Keane

Introduction

The concept of "attention" was considered to be important by many philosphers and psychologists in the late 19th century, but fell into disrepute because the behaviourists regarded all internal processes with the utmost suspicion. Attention became fashionable again following the publication of Broadbent's book *Perception and communication* in 1958, but more recently many have argued that it is too vague to be of value. Moray (1969) pointed out that attention is sometimes used to refer to the ability to select part of the incoming stimulation for further processing, but it has also been regarded as synonymous with concentration or mental set. It has been applied to search processes in which a specified target is looked for, and it has also been suggested that attention co-varies with arousal (e.g. the drowsy individual is in a state of low arousal and attends little to his or her environment).

There is an obvious danger that a concept that is used to explain everything will turn out to explain nothing. However, attention is most commonly used to refer to selectivity of processing. This was the sense emphasised by William James (1890, pp. 403–404):

> Everyone knows what attention is. It is the taking possession of the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalisation, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others.

An issue of some importance concerns the relationship between attention and consciousness. In order to discuss this, we need first to define "consciousness." According to Baars (1988, p. 15): "We will consider people to be conscious of an event if (1) they can say immediately afterwards that they were conscious of it *and* (2) we can independently verify the accuracy of their report." In the context of that definition, attention is "that which controls access to conscious experience" (Baars, 1988, p. 302). More specifically, by attending to certain visual or auditory stimuli rather than others, we can determine in part the contents of consciousness.

If we ask what makes us attend to some things rather than others, then the usual answer is that we choose to attend to sources of information that are

From chapter 5 in *Cognitive Psychology: A Student's Handbook*, 3d ed. (East Sussex, UK: Psychology Press, 1995), 95–122. Reprinted with permission.

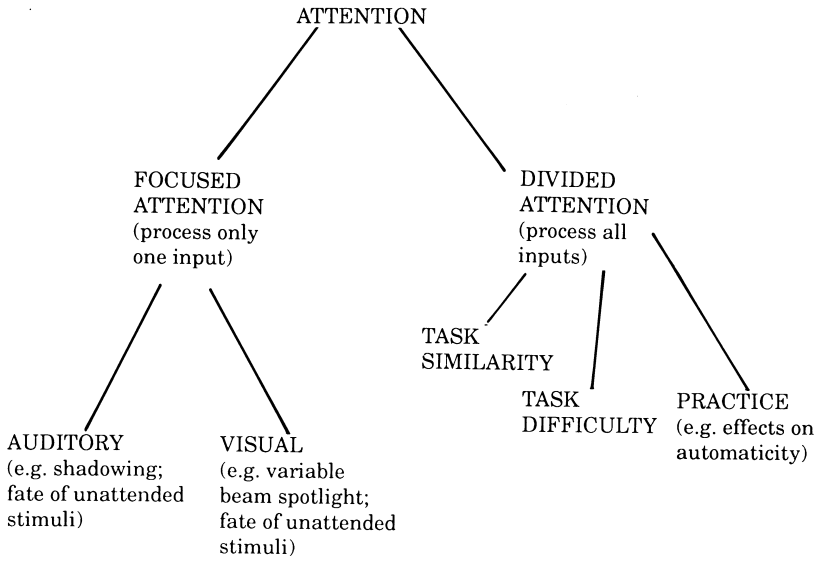


Figure 15.1
The ways in which different topics in attention are related to each other.

relevant in the context of our present activities and goals. That is true as far as it goes, but attention is sometimes "captured" involuntarily by certain stimuli. For example, Muller and Rabbitt (1989) instructed their subjects to allocate visual attention on the basis of an arrow and to ignore briefly brightened squares presented in the periphery of vision. In spite of these instructions, subjects' attention was drawn to the brightened squares.

There is an important distinction between focused and divided attention (see figure 15.1). Focused attention is studied by presenting people with two or more stimulus inputs at the same time, and instructing them to process and respond to only one. Work on focused attention can tell us how effectively people can select certain inputs rather than others, and it enables us to investigate the nature of the selection process and the fate of unattended stimuli. Divided attention is also studied by presenting at least two stimulus inputs at the same time, but with instructions that all stimulus inputs must be attended to and responded to. Studies of divided attention provide useful information about an individual's processing limitations, and may tell us something about attentional mechanisms and their capacity.

There are two important limitations in most research on attention. First, although we can attend to either the external environment or the internal environment (i.e. our own thoughts and information in long-term memory), most of the work on attention has been concerned only with attention to the external environment. Why should this be so? Experimenters can identify and control the stimuli presented in the external environment in a way that is simply not possible with internal determinants of attention.

Second, as Tipper, Lortie, and Baylis (1992) pointed out, most studies of attention are very artificial. In the real world, we generally attend to three-

dimensional people and objects, and decide what actions might be appropriate with respect to them. In the laboratory, in contrast, the emphasis, according to Tipper et al. (1992, p. 902), is on "experiments that briefly present static 2D displays and require arbitrary responses. It is clear that such experimental situations are rarely encountered in our usual interactions with the environment." Tipper et al. (1992) carried out a series of experiments under fairly naturalistic conditions. As their findings resembled those obtained in traditional laboratory studies, the artificiality of most laboratory research may not always undermine its validity.

Focused Auditory Attention

Systematic research on focused attention was initiated by the British scientist Colin Cherry (1953). He was working in an electronics research laboratory at the Massachusetts Institute of Technology, but somehow managed to find himself involved in psychological research. What fascinated Cherry was the "cocktail party" problem, i.e. how are we able to follow just one conversation when several different people are talking at once? Cherry discovered that this ability involves making use of physical differences to select among the auditory messages. These physical differences include differences in the sex of the speaker, in voice intensity, and in the location of the speaker. When Cherry presented two messages in the same voice to both ears at once (thereby eliminating these physical differences), listeners found it remarkably difficult to separate out the two messages on the basis of meaning alone.

Cherry also carried out experiments in which one auditory message had to be shadowed (i.e. repeated back, out loud) at the same time as a second auditory message was played to the other ear. Very little information seemed to be extracted from the second or non-attended message. Listeners seldom noticed when that message was spoken in a foreign language or in reversed speech. In contrast, physical changes such as the insertion of a pure tone were almost always detected. The conclusion that unattended auditory information receives practically no processing was supported by other evidence. For example, there is practically no memory for words on the unattended message even when they are presented 35 times each (Moray, 1959).

Broadbent's Theory

Broadbent (1958) felt that the findings from the shadowing task were important. He was also impressed by data from a memory task in which three pairs of digits were presented to a subject dichotically, i.e. three digits were heard one after the other by one ear, at the same time as three different digits were presented to the other ear. Subjects demonstrated a clear preference for recalling the digits ear by ear rather than pair by pair. In other words, if 496 were presented to one ear and 852 to the other ear, recall would be 496852 rather than 489562.

Broadbent (1958) accounted for the various findings by making the following assumptions (see figure 15.2):

- Two stimuli or messages presented at the same time gain access in parallel (i.e. at the same time) to a sensory buffer.

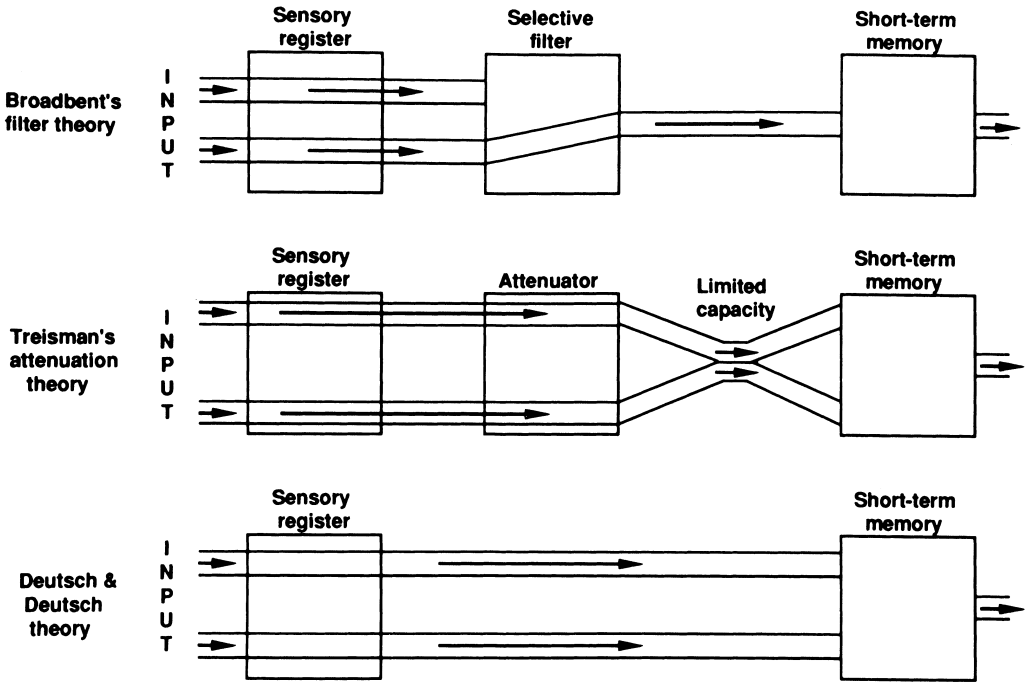


Figure 15.2
A comparison of Broadbent's theory (top): Treisman's theory (middle); and Deutsch's theory (bottom).

- One of the inputs is then allowed through a filter on the basis of its physical characteristics, with the other input remaining in the buffer for later processing.
- This filter is necessary in order to prevent overloading of the limited-capacity mechanism beyond the filter; this mechanism processes the input thoroughly.

This theory handles Cherry's basic findings, with unattended messages being rejected by the filter and thus receiving minimal processing. It also accounts for performance on Broadbent's dichotic task because the filter selects one input on the basis of the most prominent physical characteristic distinguishing the two inputs (i.e. the ear of arrival). However, it fails to explain other findings. It assumes that the unattended message is always rejected at an early stage of processing, but this is not correct. The original shadowing experiments made use of subjects who had little or no previous experience of shadowing messages, so that nearly all of their available processing resources had to be allocated to the shadowing task. Underwood (1974) asked subjects to detect digits presented on either the shadowed or the non-shadowed message. Naive subjects detected only 8% of the digits on the non-shadowed message, but an experienced researcher in the area detected 67% of the non-shadowed digits.

In most of the early work on the shadowing task, the two messages were usually rather similar (i.e. they were both auditorily presented verbal messages). Allport, Antonis, and Reynolds (1972) discovered that the degree of similarity between the two messages had a major impact on memory for the non-shadowed message. When shadowing of auditorily presented passages was combined with auditory presentation of words, memory for the words was very poor. However, when the same shadowing task was combined with picture presentation, memory for the pictures was very good (90% correct). Thus, if two inputs are dissimilar from each other, they can both be processed more thoroughly than was allowed for on Broadbent's filter theory.

In the early studies, it was concluded that there was no processing of the meaning of unattended messages because the subjects had no conscious awareness of their meaning. This left open the possibility that meaning might be processed without awareness. Von Wright, Anderson, and Stenman (1975) gave their subjects two auditorily presented lists of words, with instructions to shadow one list and to ignore the other. When a word that had previously been associated with electric shock was presented on the non-attended list, there was sometimes a noticeable physiological reaction in the form of a galvanic skin response. The same effect was produced by presenting a word very similar in sound or meaning to the shocked word. These findings suggest that information on the unattended message was processed in terms of both sound and meaning, even though the subjects were not consciously aware that the previously shocked word had been presented. However, as galvanic skin responses were detected on only a fraction of the trials, it is likely that thorough processing of unattended information occurred only some of the time.

In sum, there can be far more thorough processing of the non-shadowed message than would have been expected on Broadbent's (1958) theory. He proposed a relatively inflexible system of selective attention that cannot account for the great variability in the amount of analysis of the non-shadowed message. The same inflexibility of the filter theory is also shown in its assumption that the filter selects information on the basis of physical features. This assumption is supported by the tendency of subjects to recall dichotically presented digits ear by ear, but a small change in the basic experiment can alter the order of recall considerably. Gray and Wedderburn (1960) made use of a version of the dichotic task in which "Who 6 there" might be presented to one ear at the same time as "4 goes 1" was presented to the other ear. The preferred order of report was not ear by ear; instead, it was determined by meaning (e.g. "who goes there" followed by "4 6 1"). The implication is that selection can occur either before the processing of information from both inputs or afterwards. The fact that selection can be based on the meaning of presented information is inconsistent with filter theory.

Alternative Theories

Treisman (1964) proposed a theory in which the analysis of unattended information is attenuated or reduced (see figure 15.2). Whereas Broadbent had suggested that there was a bottleneck early in processing, Treisman claimed that the location of the bottleneck was more flexible. She proposed that stimu-

lus analysis proceeds in a systematic fashion through a hierarchy starting with analyses based on physical cues, syllabic pattern, and specific words, and moving on to analyses based on individual words, grammatical structure, and meaning. If there is insufficient processing capacity to permit full stimulus analysis, then tests towards the top of the hierarchy are omitted.

Treisman's theory accounts for the extensive processing of unattended sources of information that proved embarrassing for Broadbent, but the same facts were also explained by Deutsch and Deutsch (1963). They argued that all stimuli are fully analysed, with the most important or relevant stimulus determining the response (see figure 15.2). This theory resembles those of Broadbent and of Treisman in assuming the existence of a bottleneck in processing, but it places the bottleneck much nearer the response end of the processing system.

Treisman and Geffen (1967) provided support for Treisman's theory. Subjects shadowed one of two auditory messages, and at the same time tapped when they detected a target word in either message. According to Treisman's theory, there should be attenuated analysis of the non-shadowed message, and so fewer targets should be detected on that message than on the shadowed one. According to Deutsch and Deutsch, there is complete perceptual analysis of all stimuli, and so it might be predicted that there would be no difference in detection rates between the two messages. In fact, the shadowed or attended message showed a very large advantage in detection rates over the non-shadowed message (87% vs. 8%).

According to Deutsch and Deutsch (1967), their theory assumes that only important inputs lead to responses. As the task used by Treisman and Geffen (1967) required their subjects to make two responses (i.e. shadow and tap) to target words in the shadowed message, but only one response (i.e. tap) to targets in the non-shadowed message, the shadowed targets were more important than the non-shadowed ones.

Treisman and Riley (1969) handled this argument by carrying out a study in which exactly the same response was made to targets occurring in either message. They told their subjects to stop shadowing and to tap as soon as they detected a target in either message. Many more target words were still detected on the shadowed message than on the non-shadowed message.

Johnston and Heinz's Theory

Deutsch and Deutsch (1963) assumed that selection always occurs after full analysis of all inputs has taken place, which suggests that the processing system is rather rigid. In contrast, Johnston and Heinz (1978) proposed a more flexible model in which selection is possible at several different stages of processing. They made the following two main assumptions:

- The more stages of processing that take place prior to selection, the greater are the demands on processing capacity.
- Selection occurs as early in processing as possible given the task demands (in order to minimise demands on capacity).

Johnston and Wilson (1980) tested these theoretical ideas. Pairs of words were presented together dichotically (i.e. one word to each ear), and the task

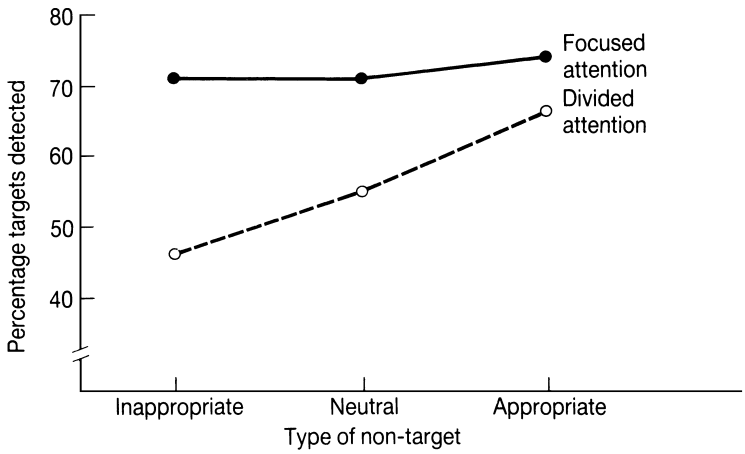


Figure 15.3
Effects of attention condition (divided vs. focused) and of type of non-target on target detection. Data from Johnston and Wilson (1980).

was to identify target items consisting of members of a designated category. The targets were ambiguous words having at least two distinct meanings. For example, if the category were "articles of clothing," then "socks" would be a possible target word. Each target word was accompanied by a non-target word biasing the appropriate meaning of the target (e.g. "smelly"), or a non-target word biasing the inappropriate meaning (e.g. "punches"), or by a neutral non-target word (e.g. "Tuesday").

When subjects did not know which ear targets would arrive at (divided attention), appropriate non-targets facilitated the detection of targets and inappropriate non-targets impaired performance (see figure 15.3). Thus, when attention needed to be divided between the two ears, there was clear evidence that the non-target words were processed for meaning. On the other hand, when subjects knew that all the targets would be presented to the left ear, the type of non-target word presented at the same time had no effect on target detection. This suggests that non-targets were not processed for meaning in this focused attention condition, and that the amount of processing received by non-target stimuli is only as much as is necessary to perform the experimental task.

Section Summary

The analysis of unattended auditory inputs can be greater than was originally thought. However, the full analysis theory of Deutsch and Deutsch (1963) seems rather dubious in view of the findings obtained by Treisman and Geffen (1967) and Treisman and Riley (1969). The most reasonable account of focused attention may be along the lines suggested by Treisman (1964), with reduced or attenuated processing of sources of information outside focal attention. The extent of such processing is probably flexible, being determined in part by task demands (Johnston & Heinz, 1978).

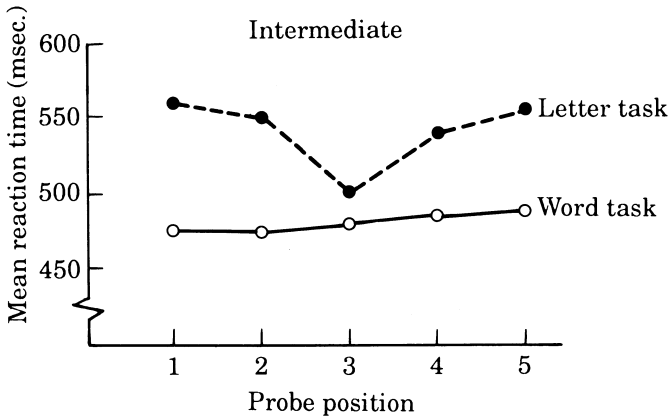


Figure 15.4
Mean reaction time to the probe as a function of probe position. The probe was presented at the time a letter string would have been presented. Data from LaBerge (1983).

Focused Visual Attention

Zoom-Lens Model

It has often been argued that focused visual attention is rather like a spotlight: everything within a relatively small area can be seen clearly, but it is much more difficult to see anything not falling within the beam of the spotlight. According to the zoom-lens model proposed by Eriksen (1990), there is an attentional spotlight, but this spotlight has an adjustable beam so that the area covered by the beam can be increased or decreased.

Relevant evidence was obtained by LaBerge (1983). In his study, five-letter words were presented. A probe requiring a rapid response was occasionally presented instead of, or immediately after, the word. The probe could appear in the spatial position of any of the five letters of the word. In one condition, an attempt was made to focus the subjects' attention on the middle letter of the five-letter word by asking them to categorise that letter. In another condition, the subjects were required to categorise the entire word. It was expected that this would lead the subjects to adopt a broader attentional beam.

The findings on speed of detection of the probe are shown in figure 15.4. In order to interpret them, we need to make the reasonable assumption that the probe was responded to faster when it fell within the central attentional beam than when it did not. On this assumption, the results indicate that the attentional spotlight can have either a very narrow (letter task) or rather broad beam (word task).

It is attractively simple to regard focused visual attention in terms of a zoom lens or variable-beam spotlight, but there is increasing evidence that the analogy is over-simplified. For example, consider a study by Juola, Bowhuis, Cooper, and Warner (1991). A target letter (L or R) which had to be identified was presented in one of three rings having the same centre: an inner, a middle, and an outer ring (see figure 15.5). The subjects fixated the centre of the display, and

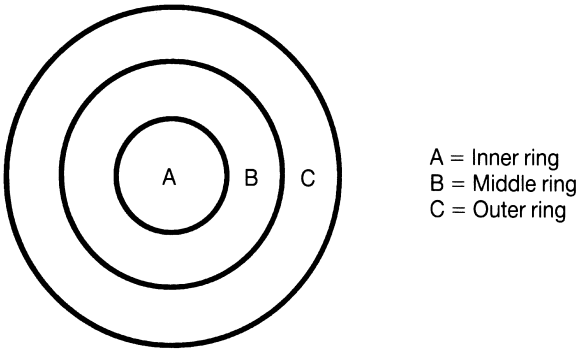


Figure 15.5
An indication of the stimulus display used by Juola et al. (1991).

were given a cue which mostly provided accurate information as to the ring in which the target would be presented. If visual attention is like a spotlight, then it would be expected that speed and accuracy of performance would be greatest for targets presented in the inner ring. In fact, performance was best when the target appeared in the ring that had been cued. This suggests that visual attention could be allocated in an O-shaped pattern to include only the outer or the middle ring.

Evidence that is even more difficult to reconcile with the zoom-lens model was reported by Neisser and Becklen (1975). They superimposed two moving scenes on top of each other, and found that their subjects could readily attend to one scene while ignoring the other. The zoom-lens model proposes that the focus of attention is a given area in visual space, but these findings suggest that this is sometimes incorrect. It appears that objects within the visual environment can be the major focus of visual attention.

Section Summary

There is some mileage in the zoom-lens model: attention is typically focused on only part of the visual environment, and the area covered by focal attention is variable. However, the evidence suggests that focused visual attention operates in a more flexible fashion than is envisaged within the zoom-lens model. Attention does not have to be focused on an entire area in visual space, but can be directed to certain objects within that area or to certain significant parts of that area.

Unattended Visual Stimuli

We saw earlier in the chapter that there is generally rather limited processing of unattended auditory stimuli. What happens to unattended visual stimuli? Johnston and Dark (1986, p. 56) reviewed the relevant evidence, and came to the following conclusion: "Stimuli outside the spatial focus of attention undergo little or no semantic processing." In contrast, Allport (1989) argued that the meaning of unattended visual stimuli is generally processed. In order to understand how these different conclusions were arrived at, it is worth considering some of the evidence.

Francolini and Egeth (1980) reported findings that are consistent with the conclusions of Johnston and Dark (1986). Subjects were presented with a circular array of red and black letters or numerals. Their task was to count the number of red items and to ignore the black items. Performance speed was reduced when the red items consisted of numerals conflicting with the answer, but there was no distraction effect from the black items. These findings suggest that there was little or no processing of the to-be-ignored black items.

Subsequent research by Driver (1989) contradicted this conclusion. He used the same task as Francolini and Egeth (1980), but focused on whether or not conflicting numerical values had been presented on the previous trial. He found that there was an interference effect, and that this interference effect was of comparable size from red and black items. The fact that performance on trial n was affected by the numerical values of distracting items presented on trial $n - 1$ means that those items must have been processed.

Driver's (1989) findings demonstrate the phenomenon of *negative priming*. In this phenomenon, the processing of a target stimulus is inhibited if that stimulus or one very similar to it was an unattended or distracting stimulus on the previous trial. For example, Tipper and Driver (1988) found that having a picture as the unattended stimulus on one trial slowed the processing of the corresponding word on the next trial. The details of the processes producing this negative priming effect are not known, but it is clear that the meaning of the unattended picture must have been processed.

Section Summary

The fact that processing and responding to attended visual stimuli are often unaffected by the nature of distracting or unattended stimuli has suggested to many theorists that there is very little processing of unattended stimuli. However, the phenomenon of negative priming indicates that this conclusion is unwarranted. It is probable that there is generally at least some processing of the meaning of unattended visual stimuli, but that this processing often does not disrupt responding to attended stimuli.

Visual Search

So far we have considered some of the general characteristics of focused visual attention. In so doing, we have not discussed in detail the various underlying processes involved in focused attention. Some progress in identifying these processes has been obtained from the use of visual search tasks. In such tasks, subjects are presented with a visual display containing a variable number of stimuli. A target stimulus (e.g. red letter G) is present on half of the trials and absent on the other half, and the subjects' task is to decide as rapidly as possible whether the target is present in the display. The effects of variations in the nature of the target and the nature of the non-targets on the speed of response are observed.

Perhaps the most influential theory based on visual search is the *feature integration theory* proposed by Treisman (1988, 1992). This theory has been criticised by various theorists including Duncan and Humphreys (1989). Duncan and Humphreys (1992) proposed an alternative explanation of the visual search

findings known as *attentional engagement theory*. Both of these theories will now be discussed.

Feature Integration Theory Treisman (1988) drew a distinction between the features of objects (e.g. colour, size, lines of particular orientation) and the objects themselves. Her theory based on this distinction includes the following assumptions:

- There is a rapid initial parallel process in which the visual features of objects in the environment are processed together; this is not dependent on attention.
- There is a second, serial process in which features are combined to form objects (e.g. a large, red chair).
- The second serial process is slower than the initial parallel process, especially when several stimuli need to be processed.
- Features can be combined by focused attending to the location of the object, in which case focused attention provides the "glue" that constructs unitary objects from the available features.
- Feature combination can also be influenced by stored knowledge (e.g. bananas are usually yellow).
- In the absence of focused attention or relevant stored knowledge, features will be combined from different objects in a random fashion, producing what are known as "illusory conjunctions."

Treisman and Gelade (1980) had previously obtained apparently good support for this feature integration theory using a visual search task. In one of their experiments, subjects searched for a target in a visual display containing between 1 and 30 items. The target was either an object (a green letter T), or it consisted of a single feature (either a blue letter or an S). When the target was a green letter T, all of the non-targets shared one feature with the target (i.e. they were either the brown letter T or the green letter X). It was predicted that focused attention would be needed to detect the former target (because it is defined by a combination of features), but that the latter target could be detected in the absence of focal attention because it is defined by a single feature.

The findings were as predicted (see figure 15.6). The number of items in the visual display had a substantial effect on detection speed when the target was defined by a combination or conjunction of features (i.e. a green letter T), presumably because focused attention was required. However, there was practically no effect of display size when the target was defined by a single feature (i.e. a blue letter or an S).

According to the feature integration theory, lack of focused attention produces a state of affairs in which the features of different objects are processed but remain "unglued." This should lead to the random combination of features and illusory conjunctions referred to earlier. This prediction was confirmed by Treisman and Schmidt (1982). They obtained numerous illusory conjunctions when attention was widely distributed, but not when the stimuli were presented to focal attention.

Treisman has modified her feature integration theory in recent years. For example, Treisman and Sato (1990) argued that the degree of similarity between

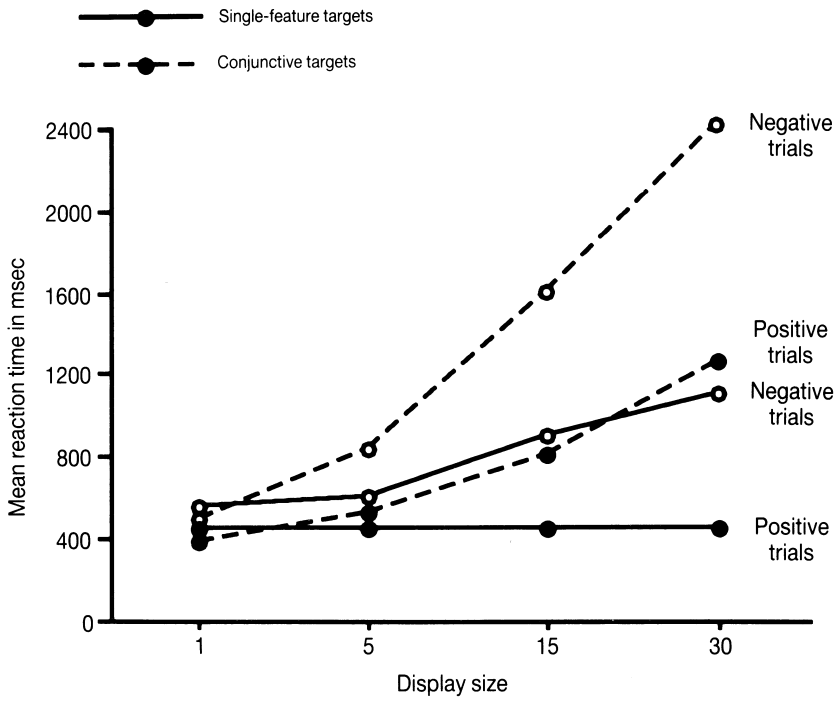


Figure 15.6
Performance speed on a detection task as a function of target definition (conjunctive vs. single feature) and display size. Adapted from Treisman and Gelade (1980).

the target and the distractors is a factor influencing visual search time. They claimed (with supporting evidence) that visual search for an object target defined by more than one feature is typically limited to those distractors possessing at least one of the features of the target. For example, if you were looking for a blue circle in a display containing blue triangles, red circles, and red triangles, then you would ignore red triangles. This contrasts with the views of Treisman and Gelade (1980), who argued that none of the stimuli would be ignored in such circumstances.

Attentional Engagement Theory Duncan and Humphreys (1989, 1992) have proposed an attentional engagement theory of visual attention. They assumed that the time taken to detect a target in a visual display depends on two major factors:

1. Search times will be slower when the similarity between the target and the non-targets is increased.
2. Search times will be slower when there is reduced similarity among non-targets. Thus, the slowest search times are obtained when non-targets are dissimilar to each other, but similar to the target.

Some evidence that visual search can be very rapid when the non-targets are all the same was obtained by Humphreys, Riddoch, and Quinlan (1985). Subjects were asked to detect a target of an inverted T against a background of

Ts the right way up. The time taken to detect the target was scarcely affected by the number of non-targets, presumably because they were all the same. According to feature integration theory, the fact that the target was defined by a combination or conjunction of features (i.e. a vertical line and a horizontal line) means that visual search should have been slow and much affected by the number of non-targets.

At a more explanatory level, the following assumptions are incorporated into the attentional engagement theory:

- There is an initial parallel stage of perceptual segmentation and analysis involving all of the visual items together.
- There is a subsequent stage of processing in which selected information is entered into visual short-term memory; this corresponds to selective attention.
- The speed of visual search depends on how easily the target item enters visual short-term memory.
- Visual items that are well matched to the description of the target item are most likely to be selected for visual short-term memory; thus, non-targets that are similar to the target slow the search process.
- Visual items that are perceptually grouped (e.g. because they are very similar) will tend to be selected (or rejected) together for visual short-term memory; thus, dissimilar non-targets cannot be rejected together and this slows the search process.

Some of the differences between this theory and Treisman's feature integration theory can be seen if we reconsider the study by Treisman and Gelade (1980). It will be remembered that there were long search times to detect a green letter T in a display containing an approximately equal number of brown Ts and green Xs (see figure 15.6), and Treisman and Gelade (1980) argued that this occurred because of the need for focal attention to produce the necessary conjunction of features. In contrast, Duncan and Humphreys (1989, 1992) claimed that the slow performance resulted from the high similarity between the target and non-target stimuli (all of the latter possessed one of the features of the target stimulus) and the dissimilarity among the non-target stimuli (the two different non-targets did not share any features).

Section Summary

The speed of visual search appears to depend on a number of factors. It is likely that the similarity between target and non-targets (accepted by Duncan and Humphreys, and by Treisman), the degree of similarity among non-targets (emphasised by Duncan and Humphreys), and conjunction of features (emphasised by Treisman) all affect visual search. There are indications that the differences between feature integration theory and attentional engagement theory are becoming less as the theories are modified. As Treisman (1992, p. 589) concluded: "There is substantial convergence between the respective theories, but it still appears that conjoining features poses a special problem that cannot be explained solely by the grouping and matching mechanisms of Duncan and Humphreys."

Disorders of Visual Attention: Cognitive Neuropsychology
Michael Posner (e.g. Posner & Petersen, 1990) proposed a theoretical framework within which various disorders of visual attention in brain-damaged patients can be understood. In essence, he argued that at least three separate abilities are involved in visual attention:

- The ability to *disengage* attention from a given visual stimulus.
- The ability to *shift* attention from one target stimulus to another.
- The ability to *engage* attention on a new visual stimulus.

Disengagment of Attention Problems with the disengagement of attention have been studied in patients suffering from what is known as *unilateral visual neglect*. The most common form of unilateral visual neglect is when patients with damage to the right hemisphere neglect or ignore visual stimuli in the left side of space. The problem is not simply one of being unable to see what is presented to the affected side because such patients can also show neglect on tasks involving images rather than visual perception (Bisiach & Luzzati, 1978).

Posner, Walker, Friedrich, and Rafal (1984) carried out a study on patients with unilateral visual neglect in which cues to the locations of forthcoming targets were presented. The patients generally coped reasonably well with this task, even when the cue and the target were both presented to the impaired visual field. However, there was one major exception: when the cue was presented to the unimpaired visual field and the target was presented to the impaired visual field, the patients' performance was extremely poor. These findings suggest that the patients found it particularly difficult to disengage their attention from visual stimuli presented to the unimpaired side of visual space.

Patients with unilateral visual neglect have suffered damage to the parietal region of the brain (Posner et al., 1984). A different kind of evidence that the parietal area is important in attention was obtained by Petersen, Corbetta, Miezin, and Shulman (1994), who made use of PET scans. They used a variety of tasks, and discovered that there was generally considerable activation within the parietal area when attention shifted from one spatial location to another.

Problems with disengaging attention are also found in patients suffering from *simultanagnosia*. In this condition, only one object (out of two or more) can be seen at any one time, even when the objects are close together in the visual field. As most of these patients have full visual fields, it seems that the attended visual object exerts a "hold" on attention that makes disengagement difficult. However, there is evidence that neglected stimuli are processed to some extent. For example, Coslett and Saffran (1991) observed strong effects of semantic relatedness between two briefly presented words in a patient with simultanagnosia.

Shifting of Attention Posner, Rafal, Choate, and Vaughan (1985) investigated problems of shifting attention by studying patients suffering from *progressive supranuclear palsy*. Such patients have damage to the midbrain. As a consequence of this brain damage, they find it very difficult to make voluntary eye movements, especially in the vertical direction. These patients were given the task of responding to visual targets, and there were sometimes cues to the loca-

tions of forthcoming targets. There was a short, intermediate, or long interval between the cue and the target. At all intervals, valid cues (i.e. cues providing accurate information about target location) speeded up responding to the targets when the targets were presented to the left or the right of the cue. However, only cues at the long interval facilitated responding when the targets were presented above or below the cues. These findings suggest that the patients had difficulty in shifting their attention in the vertical direction.

Attentional deficits apparently associated with shifting of attention have been studied in patients with *Balint's syndrome*. These patients, who have damage to the occipital-parietal area, have difficulty in reaching for stimuli using visual guidance. Humphreys and Riddoch (1993) presented two Balint's patients with 32 circles in a display; the circles were either all the same colour, or half were one colour and the other half a different colour. The circles were either close together or spaced, and the subjects' task was to decide whether they were all the same colour. On trials where there were circles of two colours, one of the patients (SA) performed much better when the circles were close together than when they were spaced (79% vs. 62%, respectively), whereas the other patient (SP) performed equivalently in the close together and spaced conditions (62% vs. 59%). Apparently some patients with Balint's syndrome (e.g. SA) find it difficult to shift attention appropriately within the visual field.

Engaging Attention Rafal and Posner (1987) investigated problems of engaging attention in patients with damage to the pulvinar nucleus of the thalamus. These patients were given the task of responding to visual targets that were preceded by cues. The patients responded faster when the cues were valid than when the cues were invalid, regardless of whether the target stimulus was presented to the same side as the brain damage or to the opposite side. However, they responded rather slowly following both kinds of cues when the target stimulus was presented to the side of the visual field opposite to that of the brain damage. According to Rafal and Posner (1987), these findings reflect a particular problem the patients have in engaging attention to such stimuli.

Additional evidence that the pulvinar nucleus of the thalamus is involved in controlling focused attention was obtained by LaBerge and Buchsbaum (1990). They took positron emission tomography (PET) measurements during an attention task, and discovered that there was increased blood flow in the pulvinar nucleus when subjects were instructed to ignore a given stimulus. Thus, the pulvinar nucleus appears to be involved in preventing attention from being focused on an unwanted stimulus as well as in directing attention to significant stimuli.

Section Summary

As Posner and Petersen (1990, p. 28) pointed out, the findings indicate that "the parietal lobe first disengages attention from its present focus, then the midbrain area acts to move the index of attention to the area of the target, and the pulvinar nucleus is involved in reading out data from the indexed locations." At a more theoretical level, the major implication is that the attentional system is considerably more complex than has been assumed by most theorists. As Allport (1989, p. 644) expressed it, "spatial attention is a distributed function in

which many functionally differentiated structures participate, rather than a function controlled uniquely by a single centre."

Divided Attention

What happens when people try to do two things at once? The answer obviously depends on the nature of the two "things." Sometimes the attempt is successful, as when an experienced motorist drives a car and holds a conversation at the same time, or a tennis player notes the position of his or her opponent while running at speed and preparing to make a stroke. At other times, as when someone tries to rub their stomach with one hand while patting their head with the other, there can be a complete disruption of performance. In this section of the chapter, we will be concerned with some of the factors determining how well two tasks can be performed concurrently (i.e. at the same time).

Hampson (1989) made the important point that focused and divided attention are more similar in some ways than one might have imagined. Factors such as use of different modalities which facilitate focused or selective attention generally also make divided attention easier. According to Hampson (1989, p. 267), the reason for this is that "anything which minimises interference between processes, or keeps them 'further apart' will allow them to be dealt with more readily either selectively or together."

At a more theoretical level, the breakdowns of performance often found when two tasks are combined shed light on the limitations of the human information-processing system. It has been assumed by many theorists that such breakdowns reflect the limited capacity of a single multi-purpose central processor or executive that is sometimes simply referred to as "attention." Other theorists are more impressed by our apparent ability to perform two relatively complex tasks at the same time without disruption or interference. Such theorists tend to favour the notion of several specific processing resources, arguing that there will be no interference between two tasks provided that they make use of different processing resources.

More progress has been made at the empirical level than at the theoretical level. It is possible to predict reasonably accurately whether or not two tasks can be combined successfully, but the accounts offered by different theorists are very diverse. Accordingly, we will make a start by discussing some of the factual evidence before moving on to the murkier issue of how the data are to be explained.

Factors Determining Dual-Task Performance

Task Similarity When we think of pairs of activities that are performed well together in everyday life, the examples that come to mind usually involve two rather dissimilar activities (e.g. driving and talking; reading and listening to music). There is much evidence that the degree of similarity between two tasks is of great importance. As we saw earlier in the chapter, when people attempt to shadow or repeat back prose passages while learning auditorily presented words, their subsequent recognition-memory performance for the words is at chance level (Allport et al., 1972). However, the same authors found that memory was excellent when the to-be-remembered material consisted of pictures.

There are various kinds of similarity that need to be distinguished. Wickens (1984) reviewed the evidence and concluded that two tasks interfere to the extent that they have the same stimulus modality (e.g. visual or auditory), make use of the same stages of processing (input, internal processing, and output), and rely on related memory codes (e.g. verbal or visual). Response similarity is also important. McLeod (1977) required subjects to perform a continuous tracking task with manual responding at the same time as a tone-identification task. Some of the subjects responded vocally to the tones, whereas others responded with the hand not involved in the tracking task. Performance on the tracking task was worse with high response similarity (manual responses on both tasks) than with low response similarity (manual responses on one task and vocal ones on the other).

Similarity of stimulus modality has probably been investigated most thoroughly. For example, Treisman and Davies (1973) found that two monitoring tasks interfered with each other much more when the stimuli on both tasks were presented in the same sense modality (visual or auditory) than when they were presented in different modalities.

Although it is clear that the extent to which two tasks interfere with each other is a function of their similarity, it is often very difficult to measure similarity. How similar are piano playing and poetry writing, or driving a car and watching a football match? Only when there is a better understanding of the processes involved in the performance of such tasks will sensible answers be forthcoming.

Practice Common sense suggests that the old saying, "Practice makes perfect," is especially applicable to dual-task performance. For example, learner drivers find it almost impossible to drive and to hold a conversation at the same time, whereas expert drivers find it relatively easy. Support for this commonsensical position was obtained by Spelke, Hirst, and Neisser (1976) in a study on two students called Diane and John. These students received five hours' training a week for four months on a variety of tasks. Their first task was to read short stories for comprehension at the same time as they wrote down words to dictation. They found this very difficult initially, and their reading speed and handwriting both suffered considerably. After six weeks of training, however, they were able to read as rapidly and with as much comprehension when taking dictation as when only reading, and the quality of their handwriting had also improved.

In spite of this impressive dual-task performance, Spelke et al. were still not satisfied. They discovered that Diane and John could recall only 35 out of the thousands of words they had written down at dictation. Even when 20 successive dictated words formed a sentence or came from a single semantic category, the two subjects were unaware of the fact. With further training, however, they learned to write down the names of the categories to which the dictated words belonged while maintaining normal reading speed and comprehension.

Spelke et al. (1976) wondered whether the popular notion that we have limited processing capacity is accurate, basing themselves on the dramatic findings with John and Diane. They observed (1976, p. 229): "People's ability to develop skills in specialised situations is so great that it may never be possible to define general limits on cognitive capacity." However, there are alternative ways of

interpreting their findings. Perhaps the dictation task was performed rather automatically, and so placed few demands on cognitive capacity, or there might have been a rapid alternation of attention between reading and writing. Hirst et al. (1980) claimed that writing to dictation was not done automatically because the subjects understood what they were writing. They also claimed that reading and dictation could only be performed together with success by the strategy of alternation of attention if the reading material were simple and highly redundant. However, they discovered that most subjects were still able to read and take dictation effectively when less redundant reading matter was used.

It is sometimes claimed that the studies by Spelke et al. (1976) and by Hirst et al. (1980) demonstrate that two complex tasks can be performed together without disruption, but this is not so. One of the subjects used by Hirst et al. was tested at dictation without reading, and made fewer than half the number of errors that occurred when reading at the same time. Furthermore, the reading task gave the subjects much flexibility in terms of when they attended to the reading matter, and such flexibility means that there may well have some alternation of attention between tasks.

There are other cases of apparently successful performance of two complex tasks, but the requisite skills were always highly practised. Expert pianists can play from seen music while repeating back or shadowing heard speech (Allport et al., 1972), and an expert typist can type and shadow at the same time (Shaffer, 1975). These studies are often regarded as providing evidence of completely successful task combination, but there are signs of interference when the data are inspected closely (Broadbent, 1982).

There are several reasons why practice might facilitate dual-task performance. First, subjects may develop new strategies for performing each of the tasks so as to minimise task interference. Second, the demands that a task makes on attentional or other central resources may be reduced as a function of practice. Third, although a task initially requires the use of several specific processing resources, practice may permit a more economical mode of functioning relying on fewer resources. These possibilities are considered in more detail a little later in the chapter.

Task Difficulty The ability to perform two tasks together undoubtedly depends on their difficulty, but there are several ways in which one task can be more difficult than another one. However, there are several studies showing the expected pattern of results. For example, Sullivan (1976) gave her subjects the two tasks of shadowing an auditory message and detecting target words on a non-shadowed message. When the shadowing task was made more difficult by using a less redundant message, fewer targets were detected on the non-shadowed message.

It has sometimes been assumed that the demands for resources of two tasks when performed together equal the sum of the demands of the two tasks when performed separately. However, the necessity to perform two tasks together often introduces fresh demands of co-ordination and avoidance of interference. Duncan (1979) asked his subjects to respond to closely successive stimuli, one requiring a left-hand response and the other a right-hand response. The relationship between each stimulus and response was either corresponding (i.e.

rightmost stimulus calling for response of the rightmost finger) or crossed (e.g. leftmost stimulus calling for response of the rightmost finger). Performance was rather poor when the relationship between stimulus and response was corresponding for one stimulus but crossed for the other. In these circumstances, the subjects were sometimes confused, as indicated by the fact that the errors were largely those expected if the inappropriate stimulus–response relationship had been selected. Thus, the uncertainty caused by mixing two different stimulus–response relationships added a complexity to performance that did not exist when only one of the tasks was performed.

Theoretical Accounts of Dual-Task Performance

Several theories of dual-task performance have been proposed over the years, and some of the main theoretical approaches are discussed here. As we will see, there have been theoretical disagreements about the relative importance of general and specific processes in this area. However, we will first of all consider the work of Welford (1952), who provided one of the first systematic attempts to account for dual-task performance.

Bottleneck Theories Welford (1952) argued that there is a bottleneck in the processing system which makes it difficult (or impossible) for two decisions about the appropriate responses for two different stimuli to be made at the same time. Much of the supporting evidence for this theory came from studies of the *psychological refractory period*. In the standard task, there are two stimuli (e.g. two lights) and two responses (e.g. button presses), and the subject's task is to respond to each stimulus as rapidly as possible. When the second stimulus is presented very shortly after the first stimulus, there is generally a marked slowing of the response to the second stimulus: this is known as the psychological refractory period effect (see Welford, 1952).

Although the existence of this psychological refractory period effect is consistent with the notion of a bottleneck in processing, it could be argued that it occurs because people are not used to having to respond to two immediately successive stimuli. However, Pashler (1993) discussed one of his experiments in which the effect was still observable after more than 10,000 trials of practice. Another objection to the notion that the delay in responding to the second stimulus reflects a bottleneck in processing is that the effect may instead be due to similarity of stimuli and/or similarity of responses.

Pashler (1990) carried out a study to decide between the bottleneck and similarity-based accounts of the psychological refractory period effect. According to the bottleneck theory, the effect should be present even when the two stimuli and the two responses differ considerably. In contrast, the effect should disappear if similarity is crucial to its existence. In one of Pashler's (1990) experiments, the stimuli were a tone requiring a vocal response and a visual letter requiring a button-push response. Some of the subjects were told the order in which the stimuli would be presented, whereas others were not. The findings are shown in figure 15.7. In spite of a lack of either stimulus or response similarity, there was a psychological refractory period effect, and it was somewhat greater when the order of the stimuli was known than when it was not. Thus, the findings provided strong support for the bottleneck position.

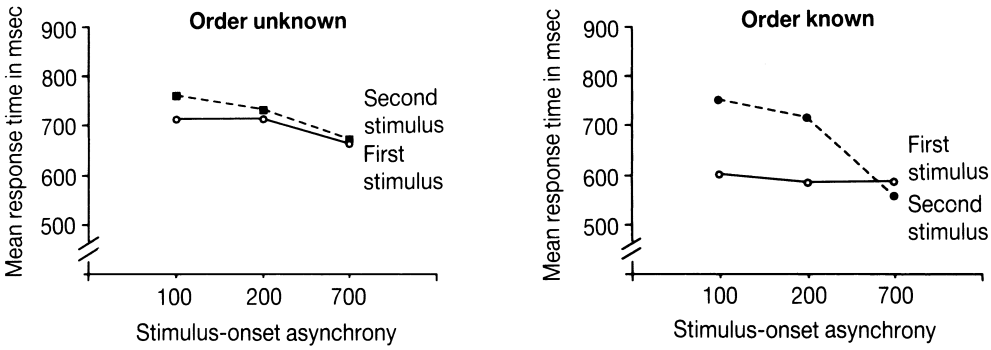


Figure 15.7
Response times to the first and second stimuli as a function of time between the onset of the stimuli (stimulus–onset asynchrony) and whether or not the order of the stimuli was known beforehand. Adapted from Pashler (1990).

Earlier in the chapter we considered various studies (e.g. Hirst et al., 1980; Spelke et al., 1976) in which two complex tasks were performed remarkably well together. Such findings make it difficult to argue for the existence of a bottleneck in processing. However, as Pashler (1993) pointed out, studies of the psychological refractory period effect have the considerable advantage that there is very precise assessment of the time taken to respond to any given stimulus. The coarse-grained measures obtained in studies such as those of Spelke et al. (1976) and Hirst et al. (1980) may simply be too insensitive to permit detection of bottlenecks.

Even if there is a bottleneck that disrupts dual-task performance, it is clearly not the only relevant factor. Accordingly, we now turn to theoretical accounts that consider other factors such as the effects of practice and similarity.

Central Capacity Theories An apparently straightforward way of accounting for many of the dual-task findings is to assume there is some central capacity which can be used flexibly across a wide range of activities (e.g. Johnston & Heinz, 1978). This central processor possesses strictly limited resources, and is sometimes known as attention or effort. The extent to which two tasks can be performed together depends on the demands that each task makes on those resources. If the combined demands of the two tasks do not exceed the total resources of the central capacity, then the two tasks will not interfere with each other. However, if the resources are insufficient to meet the demands placed on them by the two tasks, then performance disruption is inevitable.

According to central capacity theories, the crucial determinant of dual-task performance is the difficulty level of the two tasks, with difficulty being defined in terms of the demands placed on the resources of the central capacity. However, the effects of task difficulty are often swamped by those of similarity between the tasks. For example, Segal and Fusella (1970) combined image construction (visual or auditory) with signal detection (visual or auditory). As can be seen in figure 15.8, the auditory image task impaired detection of auditory signals more than the visual task did, suggesting to central capacity theorists that the auditory image task is more demanding than the visual image task. However, the auditory image task was less disruptive than the visual image

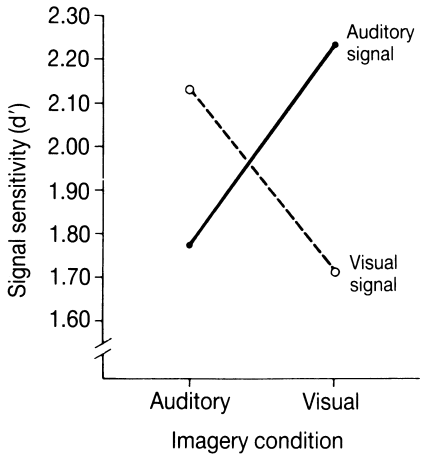


Figure 15.8
Sensitivity (d′) to auditory and visual signals as a function of concurrent imager modality (auditory vs. visual). Adapted from Segal and Fusella (1970).

task when each task was combined with a task requiring detection of visual signals, which suggests exactly the opposite conclusion. In this study, task similarity was clearly a much more important factor than task difficulty.

Many theorists have become so disenchanted with the notion of a central capacity or attentional system that they deny the existence of any such capacity or system. For example, Allport (1989, p. 647) argued that the findings "point to a multiplicity of attentional functions, dependent on a multiplicity of specialised subsystems. No one of these subsystems appears uniquely 'central.'" According to Allport, it is possible to "explain" dual-task interference by assuming that the resources of some central capacity have been exceeded, and to account for a lack of interference by assuming that the two tasks did not exceed those resources. However, in the absence of any independent assessment of central processing capacity, this is more like a re-description of the findings rather than a proper explanation.

Modular Theories The views of central capacity theorists can be compared with those of cognitive neuropsychologists. Cognitive neuropsychologists assume that the processing system is modular (i.e. it consists of numerous relatively independent processors or modules). Some of the most convincing evidence for modularity comes from the study of language in brain-damaged patients. This has revealed, for example, that reading is a complex skill involving several rather separate processing mechanisms. If the processing system consists of a large number of specific processing mechanisms, then it is clear why the degree of similarity between two tasks is so important: similar tasks compete for the same specific processing mechanisms or modules, and thus produce interference, whereas dissimilar tasks involve different modules, and so do not interfere with each other.

Allport (1989) and others have argued that dual-task performance can be accounted for in terms of modules or specific processing resources, but there are significant problems with this theoretical approach. First, there is no con-

sensus regarding the nature or number of these processing modules. Second, and following on from the first point, modular theories cannot at present be falsified. Whatever the findings of any given experiment, it is always possible to account for them after the event by postulating the existence of appropriate specific modules. Third, if there were a substantial number of modules operating in parallel, then there would be substantial problems in terms of co-ordinating their outputs in order to produce coherent behaviour.

Synthesis Theories Other theorists (e.g. Baddeley, 1986; Eysenck, 1982) have opted for a compromise position based on a hierarchical structure. The central processor, central executive, or attention is at the top of the hierarchy, and is involved in the co-ordination and control of behaviour. Below this level are specific processing mechanisms operating relatively independently of each other. It is assumed that control of these specific processing mechanisms by the central processor prevents chaos from developing.

Perhaps the major problem with the notion that there are several specific processing mechanisms and one general processing mechanism is that there does not appear to be a unitary attentional system. As we saw in the earlier discussion of cognitive neuropsychological findings, it appears that somewhat separate mechanisms are involved in disengaging, shifting, and engaging attention. If there is no general processing mechanism, then it may be unrealistic to assume that the processing system possesses a hierarchical structure.

Automatic Processing

As we saw earlier in the chapter, one of the key phenomena in studies of divided attention is the dramatic improvement that practice often has on performance. The commonest explanation for this phenomenon is that some processing activities become automatic as a result of prolonged practice. Numerous definitions of "automaticity" have been been proposed, but there is reasonable agreement on some criteria:

- Automatic processes are fast.
- Automatic processes do not reduce the capacity for performing other tasks (i.e. they demand zero attention).
- Automatic processes are unavailable to consciousness.
- Automatic processes are unavoidable (i.e. they always occur when an appropriate stimulus is presented, even if that stimulus is outside the field of attention).

As Hampson (1989, p. 264) pointed out, "Criteria for automatic processes are easy to find, but hard to satisfy empirically." For example, the requirement that automatic processes should not need attention means that they should have no influence on the concurrent performance of an attention-demanding task. This is rarely the case in practice (see Hampson, 1989, for a review). There are also problems with the unavoidability criterion. The Stroop effect, in which the naming of the colours in which words are printed is slowed down by using colour words (e.g. the word *yellow* printed in red), has often been regarded as involving unavoidable and automatic processing of the colour words. However, Kahneman and Henik (1979) discovered that the Stroop effect was much

larger when the distracting information (i.e. the colour name) was in the same location as the to-be-named colour rather than in an adjacent location. This means that the processes producing the Stroop effect are not entirely unavoidable, and thus are not completely automatic in the strict sense of the term.

Relatively few processes are fully automatic in the sense of conforming to the criteria described earlier, with a much larger number of processes being only partially automatic. Later in this section we consider a theoretical approach (that of Norman & Shallice, 1986) which distinguishes between fully automatic and partially automatic processes.

Shiffrin and Schneider's Theory

Shiffrin and Schneider (1977) and Schneider and Shiffrin (1977) argued for a theoretical distinction between controlled and automatic processes. According to them:

- Controlled processes are of limited capacity, require attention, and can be used flexibly in changing circumstances.
- Automatic processes suffer no capacity limitations, do not require attention, and are very difficult to modify once they have been learned.

Schneider and Shiffrin tested these ideas in a series of experiments. They made use of a task in which subjects memorised one, two, three, or four letters (the memory set), were then shown a visual display containing one, two, three, or four letters, and finally decided as rapidly as possible whether any one of the items in the visual display was the same as any one of the items in the memory set. In many of their experiments, the crucial manipulation was the kind of mapping used. With consistent mapping, only consonants were used as members of the memory set, and only numbers were used as distractors in the visual display (or vice versa). In other words, if a subject were given only consonants to memorise, then he or she would know that any consonant detected in the visual display must be an item from the memory set. With varied mapping, a mixture of numbers and consonants was used to form the memory set and to provide distractors in the visual display.

There were striking effects of the mapping manipulation (see figure 15.9). The numbers of items in the memory set and visual display both greatly affected decision speed in the varied mapping conditions, whereas decision speed was almost unaffected by the sizes of the memory set and visual display in the consistent mapping conditions. According to Schneider and Shiffrin (1977), a controlled search process was used with varied mapping; this involves serial comparisons between each item in the memory set and each item in the visual display until a match is achieved or until all the possible comparisons have been made. In contrast, performance with consistent mapping reflects the use of automatic processes operating independently and in parallel. According to Schneider and Shiffrin (1977), these automatic processes evolve as a result of years of practice in distinguishing between letters and numbers.

The notion that automatic processes develop through practice was tested by Shiffrin and Schneider (1977). They used consistent mapping with the consonants *b* to *l* forming one set and the consonants *q* to *z* forming the other set. As before, items from only one set were always used in the construction of the

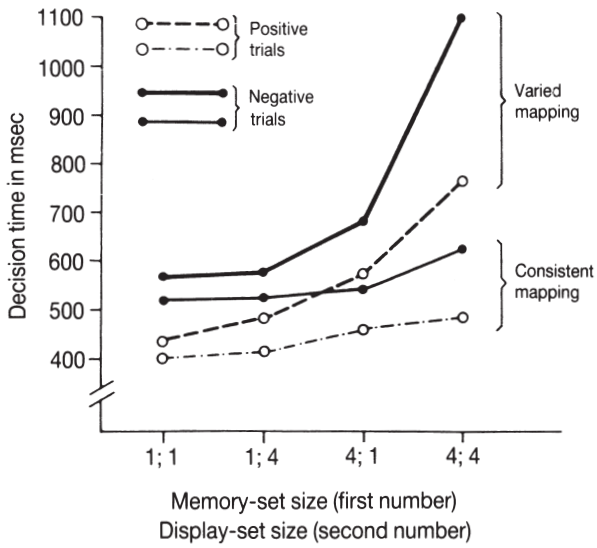


Figure 15.9
Response times on a decision task as a function of memory-set size, display-set size, and consistent versus varied mapping. Data from Shiffrin and Schneider (1977).

memory set, and the distractors in the visual display were all selected from the other set. There was a substantial improvement in performance over a total of 2100 trials, and it appeared to reflect the growth of automatic processes.

The most obvious problem with automatic processes is their lack of flexibility, which is likely to disrupt performance when there is a change in the prevailing circumstances. This was confirmed in the second part of the study just described. The initial 2100 trials with one consistent mapping were followed by a further 2400 trials with the reverse consistent mapping. This reversal of the mapping conditions had a markedly adverse effect on performance; indeed, it took nearly 1000 trials under the new conditions before performance recovered to its level at the very start of the experiment!

Shiffrin and Schneider (1977) conducted further experiments in which subjects initially attempted to locate target letters anywhere in a visual display, but were then instructed to detect targets in one part of the display and to ignore targets elsewhere in the display. Subjects were less able to ignore part of the visual display when they had developed automatic processes than when they had made use of controlled search processes. In general terms, as Eysenck (1982, p. 22) pointed out: "Automatic processes function rapidly and in parallel but suffer from inflexibility; controlled processes are flexible and versatile but operate relatively slowly and in a serial fashion."

Evaluation Shiffrin and Schneider's (1977) theoretical approach is important, but it is open to various criticisms. For example, there is a puzzling discrepancy between theory and data with respect to the identification of automaticity. The theoretical assumption that automatic processes operate in parallel and place no demands on capacity means that there should be a slope of zero (i.e. a hori-

zontal line) in the function relating decision speed to the number of items in the memory set and/or in the visual display when automatic processes are used. In fact, as can be seen in figure 15.8, decision speed was slower when the memory set and the visual display both contained several items.

The greatest weakness of Shiffrin and Schneider's approach is that it is descriptive rather than explanatory. The claim that some processes become automatic with practice is uninformative about what is actually happening. Practice may simply lead to a speeding up of the processes involved in performing a task, or it may lead to a dramatic change in the nature of the processes themselves. Cheng (1985) used the term "restructuring" to refer to the latter state of affairs. For example, if you are asked to add ten twos, you could do this in a rather laborious way by adding two and two, and then two to four, and so on. Alternatively, you could short-circuit the whole process by simply multiplying ten by two. The crucial point is that simply discovering that practice leads to automaticity does not make it clear whether the same processes are being performed more efficiently or whether entirely new processes are being used.

Cheng (1985) argued that most of Shiffrin and Schneider's findings on automaticity were actually based on restructuring. More specifically, she claimed that subjects in the consistent mapping conditions did not really search systematically through the memory set and the visual display looking for a match. If, for example, they knew that any consonant in the visual display had to be an item from the memory set, then they could simply scan the visual display looking for a consonant without any regard to which consonants were actually in the memory set.

Schneider and Shiffrin (1985) admitted that some of their earlier findings could be accounted for by assuming that subjects in consistent mapping conditions made use of knowledge about the categories being used. However, they pointed that other findings could not be explained in terms of restructuring. For example, the finding that subjects could not ignore part of the visual display after automatic processes had been acquired does not lend itself to a restructuring explanation.

Norman and Shallice's Theory

Norman and Shallice (1986) discussed a theory taking account of the distinction between fully automatic and partially automatic processes. Instead of the usual distinction between automatic and attentional or controlled processes, they identified three different levels of functioning:

- Fully automatic processing controlled by schemas (organised plans).
- Partially automatic processing involving contention scheduling without deliberate direction or conscious control; contention scheduling is used to resolve conflicts among schemas.
- Deliberate control by a supervisory attentional system.

According to Norman and Shallice (1986), fully automatic processes occur with very little conscious awareness of the processes involved. Such automatic processes would frequently disrupt behaviour if left entirely to their own devices. As a consequence, there is an automatic conflict resolution process

known as contention scheduling, which selects one of the available schemas on the basis of environmental information and current priorities. There is generally more conscious awareness of the partially automatic processes involving contention scheduling than of fully automatic processes. Finally, there is a higher level control mechanism known as the supervisory attentional system. This system is involved in decision making and trouble-shooting, and it permits flexible responding in novel situations. The supervisory attentional system may well be located in the frontal lobes.

Section Summary

The theoretical approach of Norman and Shallice (1986) incorporates the interesting notion that there are two separate control systems: contention scheduling and the supervisory attentional system. This contrasts with the views of many previous theorists that there is a single control system. The approach of Norman and Shallice is preferable, because it provides a more natural explanation for the fact that some processes are fully automatic whereas others are only partially automatic.

Automaticity as Memory Retrieval

Logan (1988) pointed out that most theories of automaticity do not indicate clearly how automaticity develops through prolonged practice. He tried to fill this gap by making these assumptions:

- Separate memory traces are stored away each time a stimulus is encountered and processed.
- Practice with the same stimulus leads to the storage of increased information about the stimulus, and about what to do with it.
- This increase in the knowledge base with practice permits rapid retrieval of relevant information when the appropriate stimulus is presented.
- "Automaticity is memory retrieval: performance is automatic when it is based on a single-step direct-access retrieval of past solutions from memory" (Logan, 1988, p. 493).
- In the absence of practice, responding to a stimulus requires thought and the application of rules; after prolonged practice, the appropriate response is stored in memory and can be accessed very rapidly.

These theoretical views make coherent sense of many of the characteristics of automaticity. Automatic processes are fast because they require only the retrieval of "past solutions" from long-term memory. Automatic processes have little or no effect on the processing capacity available to perform other tasks because the retrieval of heavily over-learned information is relatively effortless. Finally, there is no conscious awareness of automatic processes because no significant processes intervene between the presentation of a stimulus and the retrieval of the appropriate response.

In sum, Logan (1988, p. 5l9) encapsulated his theoretical position in the following way: "Novice performance is limited by a lack of knowledge rather than by a lack of resources.... Only the knowledge base changes with practice." Logan is probably right in his basic assumption that an understanding

of automatic, expert performance will require detailed consideration of the knowledge acquired with practice, rather than simply the changes in processing which occur.

Action Slips

Some of the theoretical notions considered so far in this chapter are relevant to an understanding of action slips (the performance of actions that were not intended). At the most general level, it seems clear that attentional failures usually underlie action slips, and this is recognised at a commonsensical level in the notion of "absent-mindedness." However, there are several different kinds of action slips, and each one may require its own detailed explanation.

Diary Studies

One of the main ways of studying action slips is to collect numerous examples via diary studies. Sellen and Norman (1992, p. 317) gave the following examples of action slips from a diary study: "I planned to call my sister Angela but instead called Agnes (they are twins). What I heard myself say did not match what I was thinking," and "I wanted to turn on the radio but walked past it and put my hand on the telephone receiver instead. I went to pick up the phone and I couldn't figure out why."

In one diary study, Reason (1979) asked 35 people to keep diaries of their action slips over a two-week period. Over 400 action slips were reported, most of which belonged to five major categories. Forty percent of the slips involved *storage failures*, in which intentions and actions were either forgotten or recalled incorrectly. Reason (1979, p. 74) quoted the following example of a storage failure: "I started to pour a second kettle of boiling water into a teapot of freshly made tea. I had no recollection of having just made it."

A further 20% of the errors were *test failures* in which the progress of a planned sequence was not monitored sufficiently at crucial junctures. An illustrative test failure from one person's diary went as follows (Reason, 1979, p. 73): "I meant to get my car out, but as I passed through the back porch on my way to the garage I stopped to put on my wellington boots and gardening jacket as if to work in the garden." *Subroutine failures* accounted for a further 18% of the errors; these involved insertions, omissions, or re-orderings of the component stages in an action sequence. Reason (1979, p. 73) gave the following example of this type of error: "I sat down to do some work and before starting to write I put my hand up to my face to take my glasses off, but my fingers snapped together rather abruptly because I hadn't been wearing them in the first place."

There were relatively few examples of action slips belonging to the two remaining categories of *discrimination failures* (11%) and *programme assembly failures* (5%). The former category consisted of failures to discriminate between objects (e.g. mistaking shaving cream for toothpaste), and the latter category consisted of inappropriate combinations of actions (e.g. Reason, 1979, p. 72): "I unwrapped a sweet, put the paper in my mouth, and threw the sweet into the waste bucket."

Evaluation It would be unwise to attach much significance to the percentages of the various kinds of action slips for a number of reasons. First, the figures are based on those action slips that were detected, and we simply do not know how many cases of each kind of slips went undetected. Second, the number of occurrences of any particular kind of action slip is meaningful only when we know the number of occasions on which that kind of slip might have occurred but did not. Thus, the small number of discrimination failures may reflect either good discrimination or a relative lack of situations requiring anything approaching a fine discrimination.

Another issue is that two action slips may appear to be superficially similar, and so be categorised together, even though the underlying mechanisms are different. For example, Grudin (1983) conducted videotape analyses of substitution errors in typing involving striking the key adjacent to the intended key. Some of these substitution errors involved the correct finger moving in the wrong direction, whereas others involved an incorrect key being pressed by the finger that normally strikes it. According to Grudin (1983), the former kind of error is due to faulty execution of an action, whereas the latter is due to faulty assignment of the finger. We would need more information than is generally available in most diary studies to identify such subtle differences in underlying processes.

Laboratory Studies of Action Slips

Several techniques have been used to produce action slips in laboratory conditions. What is often done is to provide a misleading context which increases the activation of an incorrect response at the expense of the correct response. Reason (1992) discussed a study of the "oak–yolk" effect illustrating this approach. Some subjects were asked to respond as rapidly as possible to a series of questions (the most frequent answers are given):

Q: What do we call the tree that grows from acorns?
A: Oak.
Q: What do we call a funny story?
A: Joke.
Q: What sound does a frog make?
A: Croak.
Q: What is Pepsi's major competitor?
A: Coke.
Q: What is another word for cape?
A: Cloak.
Q: What do you call the white of an egg?
A: Yolk.

The correct answer to the last question is "albumen." However, 85% of these subjects gave the wrong answer because it rhymed with the answers to the previous questions. In contrast, of those subjects only asked the last question, a mere 5% responded "yolk."

Although it is possible to produce large numbers of action slips under laboratory conditions, it is not clear that such slips resemble those typically found

under naturalistic conditions. As Sellen and Norman (1992, p. 334) pointed out, many naturally occurring action slips occur:

> . . . when a person is internally preoccupied or distracted, when both the intended actions and the wrong actions are automatic, and when one is doing familiar tasks in familiar surroundings. Laboratory situations offer completely the opposite conditions. Typically, subjects are given an unfamiliar, highly contrived task to accomplish in a strange environment. Most subjects arrive motivated to perform well and . . . are not given to internal preoccupation. . . . In short, the typical laboratory environment is possibly the least likely place where we are likely to see truly spontaneous, absent-minded errors.

Theories of Action Slips

At a general level, most theorists (e.g. Reason, 1992; Sellen & Norman, 1992) have assumed that action slips occur in part because there are two modes of control:

- An automatic mode, in which motor performance is controlled by schemas or organised plans; the schema that determines performance is the strongest available one.
- A conscious control mode based on some central processor or attentional system; it can oversee and override the automatic control mode.

Each mode of control has its own advantages and disadvantages. Automatic control is fast and it permits valuable attentional resources to be devoted to other processing activities. However, automatic control is relatively inflexible, and action slips occur when there is undue reliance on this mode of control. Conscious control has the advantages that it is less prone to error than automatic control and it responds flexibly to environmental changes. However, it operates relatively slowly, and is an effortful process.

It follows from this theoretical analysis that action slips occur when an individual is in the automatic mode of control and the strongest available schema or motor programme is not appropriate. The involvement of the automatic mode of control can be seen in many of Reason's (1979) action slips. One common type of action slip involves repeating an action unnecessarily because the first action has been forgotten (e.g. attempting to start a car that has already started, or brushing one's teeth twice in quick succession). We know from studies in which listeners attend to one message and repeat it back while ignoring a second message presented at the same time, that unattended information is held very briefly and then forgotten. When the initial starting of a car or brushing one's teeth occurs in the automatic mode of control, it would be predicted that subsequent memory for what has been done should be extremely poor, and so the action would often be repeated.

Sub-routine failures occur when a number of distinct motor programmes need to be run off in turn. Although each motor programme can be carried out without use of the conscious mode of control, a switch to that mode is essential at certain points in the sequence of actions, especially when a given situation is common to two or more motor programmes, and the strongest available motor

programme is inappropriate. The person who put on his gardening clothes instead of getting the car out exemplifies the way in which strong but unplanned actions can occur in the absence of attentional control.

Schema Theory A more detailed theory was proposed by Norman (1981) and by Sellen and Norman (1992). According to them, actions are determined by hierarchically organised schemas or organised plans. The highest-level schema represents the overall intention or goal (e.g. buying a present), and the lower-level schemas correspond to the actions involved in accomplishing that intention (e.g. taking money out of the bank; taking the train to the nearest shopping centre). A schema determines action when its level of activation is sufficiently high and when the appropriate triggering conditions exist (e.g. getting into the train when it stops at the station). The activation level of schemas is determined by current intentions and by the immediate environmental situation.

According to this schema model, action slips occur for various reasons:

- Errors in the formation of an intention.
- Faulty activation of a schema, leading to activation of the wrong schema or to loss of activation in the correct schema.
- Faulty triggering of active schemas, leading to action being determined by the wrong schema.

Many of the action slips recorded by Reason (1979) can be related to this theoretical framework. For example, discrimination failures can lead to errors in the formation of an intention, and storage failures for intentions can produce faulty triggering of active schemas.

Evaluation One of the positive characteristics of recent theories is the notion that errors or action slips should not be regarded as special events produced by their own mechanisms; rather, they emerge from the interplay of conscious and automatic control, and are thus "the normal by-products of the design of the human action system" (Sellen & Norman, 1992, p. 318). On the negative side, the notion that behaviour is determined by the automatic or conscious mode of control is rather simplistic. As we saw earlier in the chapter, there are considerable doubts about the notion of automatic processing, and it is improbable that there is a unitary attentional system. More needs to be discovered about the factors determining which mode of control will dominate. It is correctly predicted by contemporary theory that action slips should occur most frequently with highly practised activities, because it is under such circumstances that the automatic mode of control has the greatest probability of being used. However, the incidence of action slips is undoubtedly much greater with actions that are perceived to be of minor importance than those regarded as very important. For example, many circus performers carry out well-practised actions, but the danger element ensures that they make minimal use of the automatic mode of control. It is not clear that recent theories are equipped to explain such phenomena.

Behavioural Efficiency

It might be argued that people would function more efficiently if they placed less reliance on relatively automatic processes and more on the central pro-

cessor. However, such an argument is suspect because automated activities can sometimes be disrupted if too much attention is paid to them. For example, it can become more difficult to walk down a steep spiral staircase if attention is paid to the leg movements involved. Moreover, Reason's diarists produced an average of only one action slip per day, which does not indicate that their usual processing strategies were ineffective. Indeed, most people seem to alternate between the automatic and attention-based modes of control very efficiently. The optimal strategy involves very frequent shifts from one mode of control to the other, and it is noteworthy that these shifts are performed with great success for the most part.

Action slips are the consequences of a failure to shift from automatic to attention-based control at the right time. Although they are theoretically important, action slips usually have a minimally disruptive effect on everyday life. However, there may be some exceptions, such as absent-minded professors who focus on their own profound inner thoughts rather than on the world around them!

Section Summary

Action slips (i.e. the performance of actions that were not intended) have been investigated by means of diary studies in which subjects keep daily records of any slips they make. Various categories of action slip have been identified, but they all typically involve highly practised activities. Highly practised skills mostly do not require detailed attentional monitoring except at critical decision points. Failures of attention at such decision points cause many action slips. Failure to remember what was done a few seconds previously is responsible for many other action slips.

Evaluation of Theories of Attention

Attention: Unitary or Multiple Systems?

Most research has been based on the notion that there is a single, limited-capacity, attentional system. So far as focused attention tasks are concerned, the limitations of this system allegedly produce bottlenecks in processing. So far as divided attention tasks are concerned, attentional limitations often prevent successful performance of two tasks together, and lead to the development of automatic processes that are not reliant on attentional capacity.

One of the reasons for the long-lasting popularity of the view that attention is unitary (i.e. there is a single system) is that it fits well with introspective evidence. It seems as if we have a single attentional system which can (in the visual modality) be directed like a variable-beam spotlight to some part of the environment. However, this view is wrong. As was discussed earlier in the chapter, Posner and Petersen (1990) have identified three separate attentional processes: disengagement of attention from a stimulus; shifting of attention from one stimulus to another; and engagement of attention on a new stimulus.

The fact that attention is not unitary has grave implications for most theory and research on attention. The notion that any given process either requires

attention or does not (i.e. is automatic) is clearly a drastic over-simplification if there are a number of different attentional processes. In similar fashion, it may not be sensible to ask whether attentional selection occurs early or late in processing if there is no unitary attentional system. In the words of Allport (1993, pp. 203–204):

> There is no one uniform function, or mental operation (in general, no one causal mechanism), to which all so-called attentional phenomena can be attributed.... It seems no more plausible that there should be one unique mechanism, or computational resource, as the causal basis of all attentional phenomena than that there should be a unitary causal basis of thought, or perception, or of any other traditional category of folk psychology.... Reference to attention (or to the central executive, or even to the anterior attention system) as an unspecified causal mechanism explains nothing.

Functions of Attention

A major limitation of most theories of attention, and the research to which they have given rise, is that the functions of attention receive little consideration. In most research, what subjects attend to is determined by the experimental instructions. In the real world, however, what we attend to is determined in large measure by our motivational states and by the goal we are currently pursuing. This point is emphasised by Allport (1989, p. 664): "What is important to recognise ... is not the location of some imaginary boundary between the provinces of attention and motivation but, to the contrary, their essential interdependence."

Concern with the functions of attention suggests that attention theorists may need to change the focus of their research. For example, Allport (1989, 1993) identified the following (relatively uninvestigated) issues as being of major importance:

- Segmentation of different parallel processing streams.
- Priority assignment among multiple goals.
- Co-ordination between sensory input and action: selection for action.

Chapter Summary

The concept of "attention" is generally used in connection with either selective processing or mental effort and concentration. Selective attention has been investigated in studies of focused attention, in which the subject's task is to respond to one stimulus (the attended stimulus) and to ignore the other stimulus (the unattended stimulus). The issue of what happens to the unattended stimulus has been investigated in the auditory and visual modalities. Studies in the auditory modality suggest there is typically some processing of unattended stimuli, with the amount of such processing varying as a function of how easy it is to discriminate between the attended and unattended stimuli. Similar findings have been obtained when focused attention has been investigated in the visual modality.

Visual attention has been compared to a spotlight with an adjustable beam and to a zoom lens. However, although such analogies are intuitively appealing, there appears to be more processing of unattended visual stimuli outside the attentional beam than would be expected. It is also the case that visual attention operates in a more flexible fashion than is implied by the zoom-lens model.

Research by cognitive neuropsychologists has indicated that the attentional system is not unitary. Attention appears to involve at least three different processes (i.e. disengagement of attention from one stimulus; shifting of attention; engagement of attention on to a new stimulus), and brain damage sometimes selectively affects one or other of these processes.

Studies of divided attention involve presenting subjects with two tasks at the same time, with instructions to perform both tasks as well as possible. At an empirical level, the main issue is to identify those factors determining whether two tasks can be performed successfully at the same time. Three of the main factors are task similarity, task difficulty, and practice. Two tasks are performed well together when they are dissimilar, when they are relatively easy, and when they are well practised. In contrast, the worst levels of performance occur when two tasks are highly similar, rather difficult, and have been practised very little.

Several theorists have argued that practice leads to automatic processing. It is generally assumed that automatic processes are fast, that they do not reduce the capacity available for other tasks, and that there is no conscious awareness of them. Logan (1988) proposed that increased knowledge about what to do with different stimuli is stored away with practice, and that automaticity occurs when this information can be retrieved very rapidly.

Absent mindedness or action slips occur as a result of attentional failure. What often happens is that an individual runs off a sequence of highly practised and over-learned motor programmes. Attentional control is not required during the time each programme is running, but is needed when there is a switch from one programme to another. Failure to attend at these choice points can lead to the wrong motor programme being activated, especially if it is stronger than the appropriate programme. As optimal performance requires very frequent shifts between the presence and absence of attentional control, it is perhaps surprising that action slips are not more prevalent.

Most theory and research on attention are limited in various ways. Many of the major issues studied in attention research become relatively meaningless when it is accepted that attention is not unitary but rather involves multiple systems. Attention is closely bound up with motivation in the real world, but this interdepence of attention and motivation is not reflected in most theories of attention.

References

Allport, D. A. (1989). Attention and performance. In G. Claxton (Ed.), *Cognitive psychology: New directions*. London: Routledge & Kegan Paul.

Allport, D. A. (1993). Attention and control: have we been asking the wrong questions? A critical review of twenty-five years. In D. E. Meyer & S. M. Kornblum (Eds.), *Attention and Performance* (Vol. XIV). London: MIT Press.

Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology, 24*, 225–235.

Baars, B. J. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Bisiach, E., & Luzzati, C. (1978). Unilateral neglect of representational space. *Cortex, 14*, 129–133.

Broadbent, D. E. (1958). *Perception and communication*. Oxford: Pergamon.

Cheng, P. W. (1985). Restructuring versus automaticity: Alternative accounts of skills acquisition. *Psychological Review, 92*, 414–423.

Cherry, E. C. (1953). Some experiments on the recognition of speech with one and two ears. *Journal of the Acoustical Society of America, 25*, 975–979.

Coslett, H. B., & Saffran, E. M. (1991). Simultanagnosia: To see but not two see. *Brain, 114*, 1523–1545.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*, 80–90.

Deutsch, J. A., & Deutsch, D. (1967). Comments on "Selective attention: Perception or response?" *Quarterly Journal of Experimental Psychology, 19*, 362–363.

Duncan, J. (1979). Divided attention: The whole is more than sum of its parts. *Journal of Experimental Psychology: Human Perception, 5*, 216–228.

Duncan, J., & Humphreys, G. W. (1989). A resemblance theory of visual search. *Psychological Review, 96*, 433–458.

Duncan, J. W., & Humphreys, G. W. (1992). Beyond the search surface: Visual search and attentional engagement. Journal of *Experimental Psychology: Human Perception and Performance, 18*, 578–588.

Eriksen, C. W. (1990). Attentional search of the visual field. In D. Brogan (Ed.), *Visual Search*. London: Taylor & Francis.

Eysenck, M. W. (1982). *Attention and arousal: Cognition and performance*. Berlin: Springer.

Francolini, C. N., & Egeth, H. E. (1980). On the non-automaticity of automatic activation: Evidence of selective seeing. *Perception and Psychophysics, 27*, 331–342.

Gray, J. A., & Wedderburn, A. A. (1960). Grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology, 12*, 180–184.

Grudin, J. T. (1983). Error patterns in novice and skilled transcription typing. In W. E. Cooper (Ed.), *Cognitive aspects of skilled typewriting*. New York: Springer.

Hampson, P. J. (1989). Aspects of attention and cognitive science. *The Irish Journal of Psychology, 10*, 261–275.

Hirst, W., Spelke, E. S., Reaves, C. C., Caharack, G., & Neisser, U. (1980). Dividing attention without alteration or automaticity. *Journal of Experimental Psychology: General, 109*, 98–117.

Humphreys, G. W., & Riddoch, M. J. (1993). Interactions between object and space systems revealed through neuropsychology. In D. E. Meyer & S. M. Kornblum (Eds.), *Attention and performance* (Vol. XIV). London: MIT Press.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1985). Interactive processes in perceptual organization: Evidence from visual agnosia. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance* (Vol. XI). Hillsdale, NJ: Erlbaum.

James, W. (1890). *Principles of psychology*. New York: Holt.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Review of Psychology, 37*, 43–75.

Johnston, W. A., & Heinz, S. P. (1978). Flexibility and capacity demands of attention. *Journal of Experimental Psychology: General, 107*, 420–435.

Johnston, W. A., & Wilson, J. (1980). Perceptual processing of non-targets in an attention task. *Memory and Cognition, 8*, 372–377.

Juola, J. F., Bowhuis, D. G., Cooper, E. E., & Warner, C. B. (1991). Control of attention around the fovea. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 315–330.

Kahneman, D., & Henik, A. (1979). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erblaum.

Laberge, D. (1983). Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 371–379.

LaBerge, D., & Buchsbaum, M. S. (1990). Positron emission tomography measurements of pulvinar activity during an attention task. *Journal of Neuroscience, 10*, 613–619.

Logan, G. D. (1988). Toward an increase theory of automatisation. *Psychological Review, 95*, 492–527.

McLeod, P. (1977). A dual-task response modality effect: Support for multiprocessor models of attention. *Quarterly Journal of Experimental Psychology, 29*, 651–667.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11*, 56–60.

Moray, N. (1969). *Attention: Selective processes in vision and hearing*. London: Hutchinson.

Muller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 125–141.

Neisser, U., & Becklen, P. (1975). Selective looking: Attending to visually superimposed events. *Cognitive Psychology, 7*, 480–494.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *The design of everyday things*. New York: Doubleday.

Pashler, H. (1990). Do response modality effects support multiprocessor models of divided attention? *Journal of Experimental Psychology: Human Perception and Performance, 16*, 826–842.

Petersen, S. E., Corbetta, M., Miezin, F. M., & Shulman, G. L. (1994). PET studies of parietal involvement in spatial attention: Comparison of different task types. *Canadian Journal of Experimental Psychology, 48*, 319–338.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25–42.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal lobe injury on covert orienting of visual attention. *Journal of Neuroscience, 4*, 1863–1874.

Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughn, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology, 2*, 211–228.

Rafal, R. D., & Posner, M. I. (1987). Deficits in human visual spatial attention following thalamic lesions. *Proceedings of the National Academy of Science, 84*, 7349–7353.

Reason, J. T. (1979). Actions not as planned: The price of automatisation. In G. Underwood & R. Stevens (Eds.), *Aspects of consciousness: Vol. I: Psychological issues*. London: Academic Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychological Review, 84*, 1–66.

Segal, S. J., & Fusella, V. (1970). Influence of imaged pictures an sounds on detection of visual and auditory signals. *Journal of Experimental Psychology, 83*, 458–464.

Sellen, A. J., & Norman, D. A. (1992). The psychology of slips. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition*. New York: Plenum Press.

Shaffer, L. H. (1975). Multiple attention in continuous verbal tasks. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance* (Vol. V). London: Academic Press.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–190.

Spelke, E. S., Hirst, W. C., & Neisser, U. (1976). Sills of divided attention. *Cognition, 4*, 215–230.

Sullivan, L. (1976). Selective attention and secondary message analysis: A reconsideration of Broadbent's filter model of selective attention. *Quarterly Journal of Experimental Psychology, 28*, 167–178.

Tipper, S. P., Lortie, C., & Baylis, G. C. (1992). Selective reaching: Evidence for the action-centred attention. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 891–905.

Treisman, A. M. (1964). Verbal cues, language, and meaning in selective attention. *American Journal of Psychology, 77*, 206–219.

Treisman, A. M. (1992). Spreading suppression or feature intergration? A reply to Duncan and Humphreys (1992). *Journal of Experimental Psychology: Human Perception and Performance, 18*, 589–593.

Treisman, A. M., & Davies, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.), *Attention and performance* (Vol. IV). London: Academic Press.

Treisman, A. M., & Geffen, G. (1967). Selective attention: Perception or response? *Quarterly Journal of Experimental Psychology, 19*, 1–18.

Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Treisman, A. M., & Riley, J. G. A. (1969). Is selective attention selective perception or selective response: A further test. *Journal of Experimental Psychology, 79*, 27–34.

Treisman, A. M., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 459–478.

Treisman, A. M., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*, 107–141.

Underwood, G. (1974). Moray vs. the rest: The effects of extended shadowing practice. *Quarterly Journal of Experimental Psychology, 26*, 368–372.

Von Wright, J. M., Anderson, K., & Stenman, U. (1975). Generalisation of conditioned G. S.Rs in dichotic listening. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance* (Vol. V). London: Academic Press.

Welford, A. T. (1952). The psychological refractory period and the timing of high speed performance. *British Journal of Psychology, 43*, 2–19.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention*. London: Academic Press.

Chapter 16

Features and Objects in Visual Processing

Anne Treisman

If you were magically deposited in an unknown city, your first impression would be of recognizable objects organized coherently in a meaningful framework. You would see buildings, people, cars, and trees. You would not be aware of detecting colors, edges, movements, and distances, and of assembling them into multidimensional wholes for which you could retrieve identities and labels from memory. In short, meaningful wholes seem to precede parts and properties, as the Gestalt psychologists emphasized many years ago.

This apparently effortless achievement, which you repeat innumerable times throughout your waking hours, is proving very difficult to understand or to simulate on a computer—much more difficult, in fact, than the understanding and simulation of tasks that most people find quite challenging, such as playing chess or solving problems in logic. The perception of meaningful wholes in the visual world apparently depends on complex operations to which a person has no conscious access, operations that can be inferred only on the basis of indirect evidence.

Nevertheless, some simple generalizations about visual information processing are beginning to emerge. One of them is a distinction between two levels of processing. Certain aspects of visual processing seem to be accomplished simultaneously (that is, for the entire visual field at once) and automatically (that is, without attention being focused on any one part of the visual field). Other aspects of visual processing seem to depend on focused attention and are done serially, or one at a time, as if a mental spotlight were being moved from one location to another.

In 1967, Ulric Neisser, then at the University of Pennsylvania, suggested that a "preattentive" level of visual processing segregates regions of a scene into figures and ground so that a subsequent, attentive level can identify particular objects. More recently, David C. Marr, investigating computer simulation of vision at the Massachusetts Institute of Technology, found it necessary to establish a "primal sketch": a first stage of processing, in which the pattern of light reaching an array of receptors is converted into a coded description of lines, spots, or edges and their locations, orientations, and colors. The representation of surfaces and volumes and finally the identification of objects could begin only after this initial coding.

In brief, a model with two or more stages is gaining acceptance among psychologists, physiologists, and computer scientists working in artificial intelligence. Its first stage might be described as the extraction of features from

From *Scientific American* 255, no. 5 (1986): 114–125. Reprinted with permission.

patterns of light; later stages are concerned with the identification of objects and their settings. The phrase "features and objects" is therefore a three-word characterization of the emerging hypothesis about the early stages of vision.

I think there are many reasons to agree that vision indeed applies specialized analyzers to decompose stimuli into parts and properties, and that extra operations are needed to specify their recombination into the correct wholes. In part the evidence is physiological and anatomical. In particular, the effort to trace what happens to sensory data suggests that the data are processed in different areas of considerable specialization. One area concerns itself mainly with the orientation of lines and edges, another with color, still another with directions of movement. Only after processing in these areas do data reach areas that appear to discriminate between complex natural objects.

Some further evidence is behavioral. For example, it seems that visual adaptation (the visual system's tendency to become unresponsive to a sustained stimulus) occurs separately for different properties of a scene. If you stare at a waterfall for a few minutes and then look at the bank of the river, the bank will appear to flow in the opposite direction. It is as if the visual detectors had selectively adapted to a particular direction of motion independent of *what* is moving. The bank looks very different from the water, but it nonetheless shows the aftereffects of the adaptation process.

How can the preattentive aspect of visual processing be further subjected to laboratory examination? One strategy is suggested by the obvious fact that in the real world parts that belong to the same object tend to share properties: they have the same color and texture, their boundaries show a continuity of lines or curves, they move together, they are at roughly the same distance from the eye. Accordingly the investigator can ask subjects to locate the boundaries between regions in various visual displays and thus can learn what properties make a boundary immediately salient—make it "pop out" of a scene. These properties are likely to be the ones the visual system normally employs in its initial task of segregating figure from ground.

It turns out that boundaries are salient between elements that differ in simple properties such as color, brightness, and line orientation but not between elements that differ in how their properties are combined or arranged (figure 16.1). For example, a region of *T*s segregates well from a region of tilted *T*s but not from a region of *L*s made of the same components as the *T*s (a horizontal line and a vertical line). By the same token, a mixture of blue *V*s and red *O*s does not segregate from a mixture of red *V*s and blue *O*s. It seems that the early "parsing" of the visual field is mediated by separate properties, not by particular combinations of properties. That is, analysis of properties and parts precedes their synthesis. And if parts or properties are identified before they are conjoined with objects, they must have some independent psychological existence.

This leads to a strong prediction, which is that errors of synthesis should sometimes take place. In other words, subjects should sometimes see illusory conjunctions of parts or properties drawn from different areas of the visual field. In certain conditions such illusions take place frequently. In one experiment my colleagues and I flashed three colored letters, say a blue *X*, a green *T*, and a red *O*, for a brief period (200 milliseconds, or a fifth of a second) and

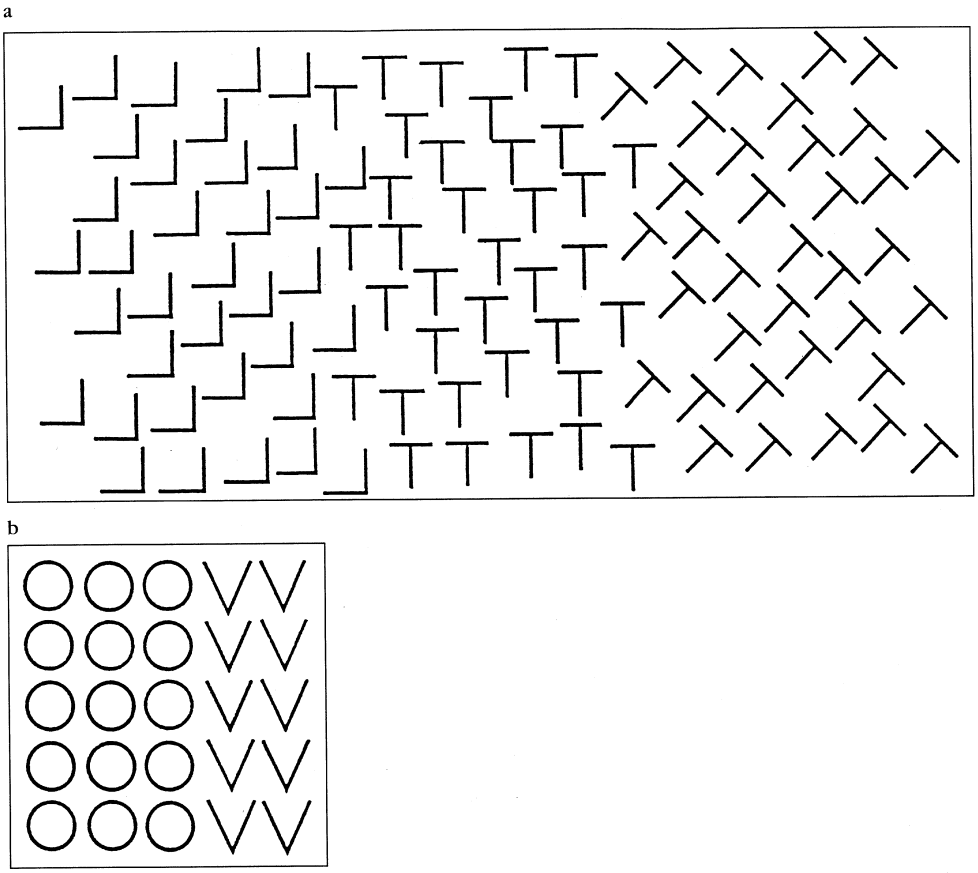


Figure 16.1
Boundaries that "pop out" of a scene are likely to reveal the simple properties, or features, of the visual world that are seized on by the initial stage of visual processing. For example, a boundary between *Ts* and tilted *Ts* pops out, whereas a boundary between *Ts* and *Ls* does not (a). The implication is that line orientations are important features in early visual processing but that particular arrangements of conjunctions of lines are not. A boundary between *Os* and *Vs* pops out (b). The implication is that simple shape properties (such as line curvature) are important.

diverted our subjects' attention by asking them to report first a digit shown at each side of the display and only then the colored letters. In about one trial in three, the subjects reported the wrong combinations—perhaps a red *X*, a green *O*, or a blue *T*.

The subjects made these conjunction errors much more often than they reported a color or shape that was not present in the display, which suggests that the errors reflect genuine exchanges of properties rather than simply misperceptions of a single object. Many of these errors appear to be real illusions, so convincing that subjects demand to see the display again to convince themselves that the errors were indeed mistakes.

We have looked for constraints on the occurrence of such illusory conjunctions. For example, we have asked whether objects must be similar for their

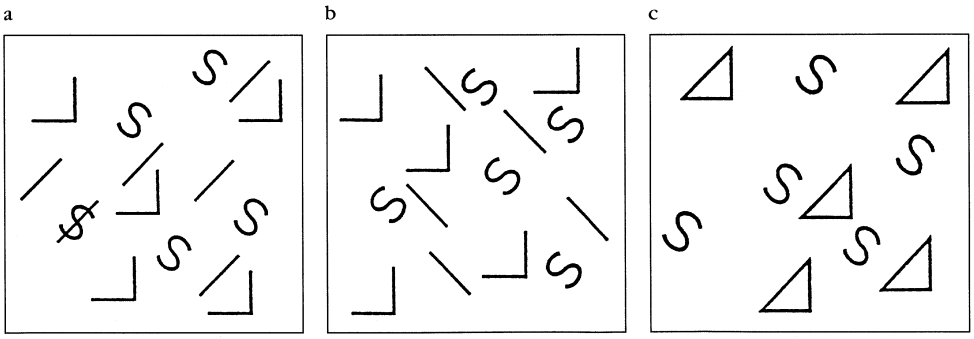


Figure 16.2
Illusory dollar signs are an instance of false conjunctions of features. Subjects were asked to look for dollar signs in the midst of *S*s and line segments (a). They often reported seeing the signs when the displays to which they were briefly exposed contained none (b). They had the same experience about as often when the line segment needed to complete a sign was embedded in a triangle (c). The experiment suggests that early visual processing can detect the presence of features independent of location.

properties to be exchanged. It seems they do not: Subjects exchanged colors between a small, red outline of a triangle and a large, solid blue circle just as readily as they exchanged colors between two small outline triangles. It is as if the red color of the triangle were represented by an abstract code for red rather than being incorporated into a kind of analogue of the triangle that also encodes the object's size and shape.

We also asked if it would be harder to create illusory conjunctions by detaching a part from a simple unitary shape, such as a triangle, than by moving a loose line. The answer again was no. Our subjects saw illusory dollar signs in a display of *S*s and lines. They also saw the illusory signs in a display of *S*s and triangles in which each triangle incorporated the line the illusion required (figure 16.2). In conscious experience the triangle looks like a cohesive whole. Nevertheless, at the preattentive level, its component lines seem to be detected independently.

To be sure, the triangle may have an additional feature, namely the fact that its constituent lines enclose an area, and this property of closure might be detected preattentively. If so, the perception of a triangle might require the detection of its three component lines in the correct orientations and also the detection of closure. We should then find that subjects do not see illusory triangles when they are given only the triangles' separate lines in the proper orientations (figure 16.3). They may need a further stimulus, a different closed shape (perhaps a circle), in order to assemble illusory triangles. That is indeed what we found.

Another way to make the early, preattentive level of visual processing the subject of laboratory investigation is to assign visual-search tasks. That is, we ask subjects to find a target item in the midst of other, "distractor" items. The assumption is that if the preattentive processing occurs automatically and across the visual field, a target that is distinct from its neighbors in its preattentive representation in the brain should "pop out" of the display. The pro-

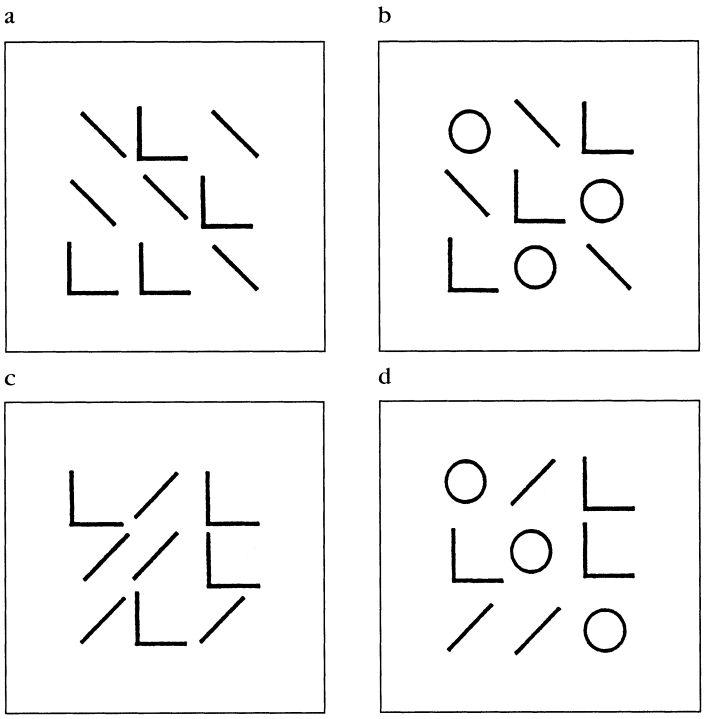


Figure 16.3
Illusory triangles constitute a test of what features must be available to support the perception of triangles. Subjects seldom reported seeing a triangle when they were briefly exposed to displays consisting of the line segments that make up a triangle (a). They saw triangles far more often when the displays also included closed stimuli, that is, shapes that enclose a space, in this case *Os* (b). Evidently, closure is a feature analyzed in early visual processing. This conclusion was supported by showing displays that lack the diagonal line to make a triangle (c, d). Subjects seldom saw triangles in such displays.

verbial needle in a haystack is hard to find because it shares properties of length, thickness and orientation with the hay in which it is hidden. A red poppy in a haystack is a much easier target; its unique color and shape are detected automatically.

We find that if a target differs from the distractors in some simple property, such as orientation or color or curvature, the target is detected about equally fast in an array of 30 items and in an array of three items. Such targets pop out of the display, so that the time it takes to find them is independent of the number of distractors. This independence holds true even when subjects are not told what the unique property of the target will be. The subjects take slightly longer overall, but the number of distractors still has little or no effect.

On the other hand, we find that if a target is characterized only by a conjunction of properties (for example, a red *O* among red *N*s and green *O*s), or if it is defined only by its particular combination of components (for example, an *R* among *P*s and *Q*s that together incorporate all the parts of the *R*), the time taken to find the target or to decide that the target is not present increases

linearly with the number of distractors. It is as if the subjects who are placed in these circumstances are forced to focus attention in turn on each item in the display in order to determine how the item's properties or parts are conjoined. In a positive trial (a trial in which a target is present) the search ends when the target is found; on the average, therefore, it ends after half of the distractors have been examined. In a negative trial (in which no target is present) all the distractors have to be checked. As distractors are added to the displays, the search time in positive trials therefore increases at half the rate of the search time in negative trials.

The difference between a search for simple features and a search for conjunctions of features could have implications in industrial settings. Quality-control inspectors might, for example, take more time to check manufactured items if the possible errors in manufacture are characterized by faulty combinations of properties than they do if the errors always result in a salient change in a single property. Similarly, each of the symbols representing, say, the destinations for baggage handled at airline terminals should be characterized by a unique combination of properties.

In a further series of experiments on visual-search tasks, we explored the effect of exchanging the target and the distractors. That is, we required subjects to find a target distinguished by the fact that it *lacks* a feature present in all the distractors. For example, we employed displays consisting of *O*s and *Q*s, so that the difference between the target and the distractors is that one is simply a circle whereas the other is a circle intersected by a line segment (figure 16.4). We found a remarkable difference in the search time depending on whether the target was the *Q* and had the line or was the *O* and lacked the line. When the target had the line, the search time was independent of the number of distractors. Evidently, the target popped out of the display. When the target lacked the line, the search time increased linearly with the number of distractors. Evidently, the items in the display were being subjected to a serial search.

The result goes against one's intuitions. After all, each case involves the same discrimination between the same two stimuli: *O*s and *Q*s. The result is consistent, however, with the idea that a pooled neural signal early in visual processing conveys the presence but not the absence of a distinctive feature. In other words, early vision extracts simple properties, and each type of property triggers activity in populations of specialized detectors. A target with a unique property is detected in the midst of distractor items simply by a check on whether the relevant detectors are active. Conversely, a target lacking a property that is present in the distractors arouses only slightly less activity than a display consisting exclusively of distractors. We propose, therefore, that early vision sets up a number of what might be called *feature maps*. They are not necessarily to be equated with the specialized visual areas that are mapped by physiologists, although the correspondence is suggestive.

We have exploited visual-search tasks to test a wide range of candidate features we thought might pop out of displays and so reveal themselves as primitives: basic elements in the language of early vision. The candidates fell into a number of categories: quantitative properties such as length or number; properties of single lines such as orientation or curvature; properties of line arrange-

ments; topological and relational properties such as the connectedness of lines, the presence of the free ends of lines or the ratio of the height to the width of a shape.

Among the quantitative candidates, my colleagues and I found that some targets popped out when their discriminability was great. In particular, the more extreme targets—the longer lines, the darker grays, the pairs of lines (when the distractors were single lines)—were easier to detect. This suggests that the visual system responds positively to "more" in these quantitative properties and that "less" is coded by default. For example, the neural activity signaling line length might increase with increasing length (up to some maximum), so that a longer target is detected against the lower level of background activity produced by short distractors. In contrast, a shorter target, with its concomitant lower rate of firing, is likely to be swamped by the greater activity produced by the longer distractors. Psychophysicists have known for more than a century that the ability to distinguish differences in intensity grows more acute with decreasing background intensity. We suggest that the same phenomenon, which is known as Weber's law, could account for our findings concerning the quantitative features.

Our tests of two simple properties of lines, orientation and curvature, yielded some surprises. In both cases we found pop-out for one target, a tilted line among vertical distractors and a curved line among straight lines, but not for the converse target, a vertical line among tilted distractors and a straight line among curves. These findings suggest that early vision encodes tilt and curvature but not verticality or straightness. That is, the vertical targets and the straight targets appear to lack a feature the distractors possess, as if they represent null values on their respective dimensions. If our interpretation is correct, it implies that in early vision, tilt and curvature are represented relationally, as deviations from a standard or norm that itself is not positively signaled.

A similar conclusion emerged for the property of closure. We asked subjects to search for complete circles in the midst of circles with gaps and for circles with gaps among complete circles. Again we found a striking asymmetry, this time suggesting that the gap is preattentively detectable but that closure is not—or rather that it becomes preattentively detectable only when the distractors have very large gaps (that is, when they are quite open shapes like semicircles). In other words, closure is preattentively detectable, but only when the distractors do not share it to any significant degree. On the other hand, gaps (or the line ends that gaps create) are found equally easily whatever their size (unless they are too small for a subject, employing peripheral vision, to see).

Finally, we found no evidence that any property of line arrangements is preattentively detectable. We tested intersections, junctions, convergent lines and parallel lines. In every case we found that search time increases with an increasing number of distractors. The targets become salient and obvious only when the subject's attention is directed to them; they do not emerge automatically when that attention is disseminated throughout the display.

In sum, it seems that only a small number of features are extracted early in visual processing. They include color, size, contrast, tilt, curvature, and line

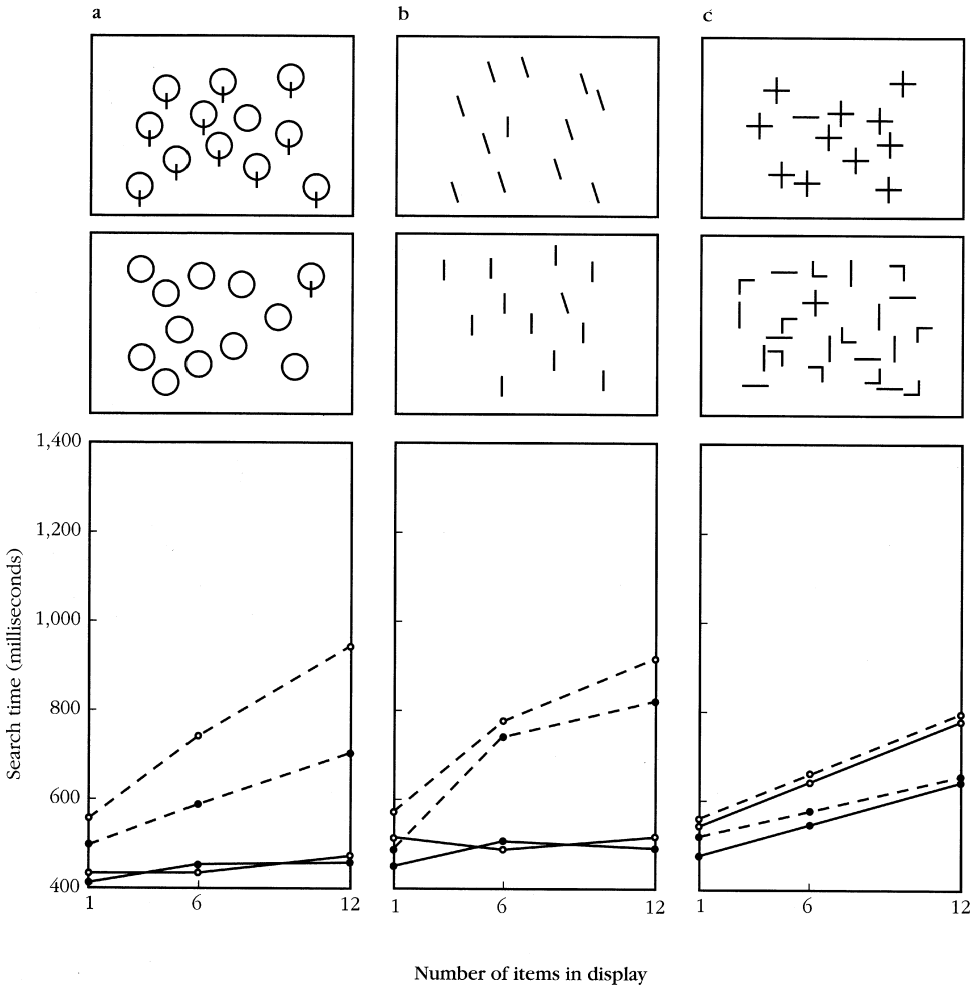


Figure 16.4
Presence of absence of a feature can have remarkably different effects on the time it takes to find a target in the midst of distractors. In one experiment (a) the target was a circle intersected by a vertical line segment or a circle without that feature. The search time for the intersected circle (solid) proved to be largely independent of the number of items in the display, suggesting that the feature popped out. The search time for the plain circle (dashed) increased steeply as distractors were added, suggesting that a serial search of the display was being made. A second experiment (b) required subjects to search for a vertical line (dashed) or a tilted line (solid). The tilted line could be found much faster; evidently only the tilted line popped out of the displays. A third experiment (c) tested an isolated line segment (dashed) or intersecting lines in the form of a plus sign (solid). Evidently neither popped out.

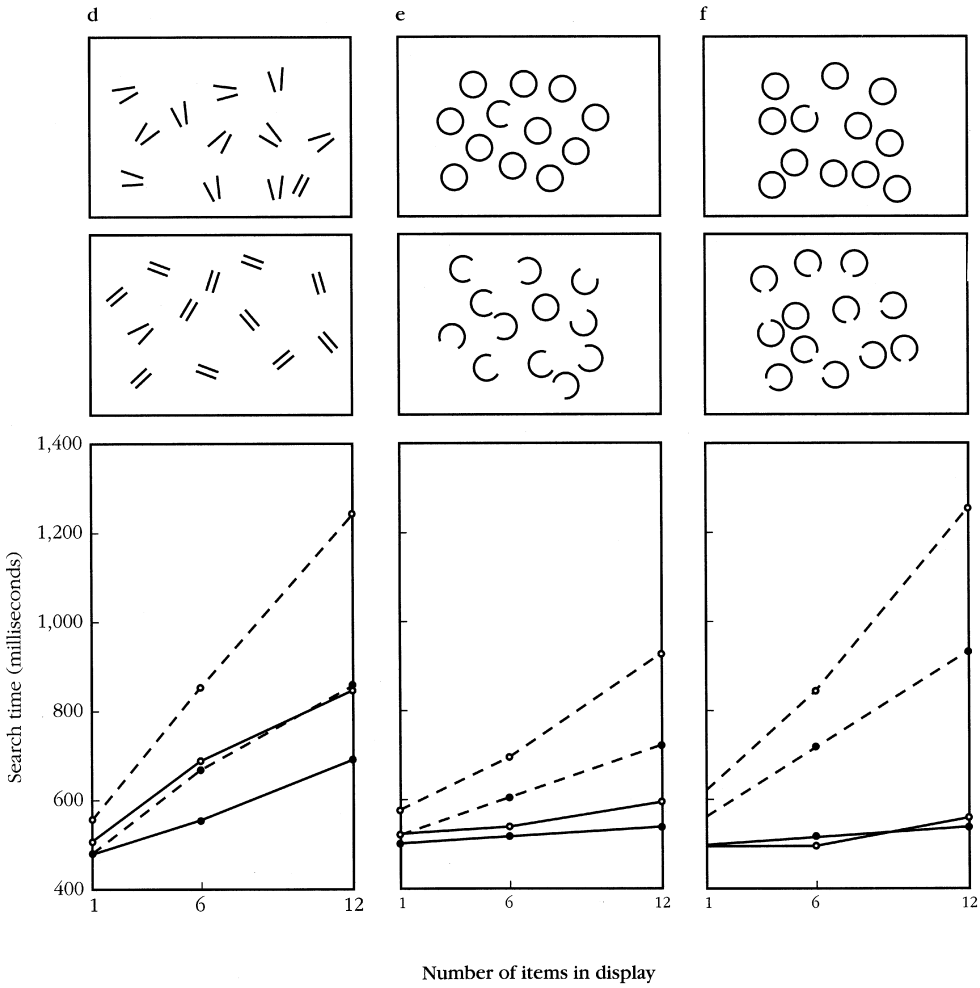


Figure 16.4
A fourth experiment (d) tested parallel lines (dashed) or converging lines (solid). Again neither popped out. A fifth experiment (e) tested closure with complete circles (dashed) or circles with a gap of a fourth of their circumference (solid). A sixth experiment (f), again testing closure, had complete circles (dashed) or circles with smaller gaps (solid). The size of the gap seemed to make no difference: The incomplete circle popped out. On the other hand, a complete circle became harder to find as the size of the gaps in distractors was reduced. Open dots represent data from trials in which the display included only distractors.

ends. Research by other investigators shows that movement and differences in stereoscopic depth are also extracted automatically in early vision. In general the building blocks of vision appear to be simple properties that characterize local elements, such as points or lines, but not the relations among them. Closure appears to be the most complex property that pops out preattentively. Finally, our findings suggest that several preattentive properties are coded as values of deviation from a null, or reference, value.

Up to this point I have concentrated on the initial, preattentive stages of vision. I turn now to the later stages. In particular I turn to the evidence that focused attention is required for conjoining the features at a given location in a scene and for establishing structured representations of objects and their relations.

One line of evidence suggesting that conjunctions require attention emerges from experiments in which we asked subjects to identify a target in a display and say where it was positioned. In one type of display only a simple feature distinguished the target from the distractors. For example, the target was a red *H* in the midst of red *O*s and blue *X*s or an orange *X* among red *O*s and blue *X*s. In other displays, the target differed only in the way its features were conjoined. For example, it was a blue *O* or a red *X* among red *O*s and blue *X*s.

We were particularly interested in the cases in which a subject identified the target correctly but gave it the wrong location. As we expected, the subjects could sometimes identify a simple target, say a target distinguished merely by its color, but get its location wrong. Conjunction targets were different: The correct identification was completely dependent on the correct localization. It does indeed seem that attention must be focused on a location in order to combine the features it contains.

In a natural scene, of course, many conjunctions of features are ruled out by prior knowledge. You seldom come across blue bananas or furry eggs. Preattentive visual processing might be called "bottom up," in that it happens automatically, without any recourse to such knowledge. Specifically, it happens without recourse to "top down" constraints. One might hypothesize that conjunction illusions in everyday life are prevented when they conflict with top-down expectations. There are many demonstrations that we do use our knowledge of the world to speed up perception and to make it more accurate. For example, Irving Biederman of the State University of New York at Buffalo asked subjects to find a target object such as a bicycle in a photograph of a natural scene or in a jumbled image in which different areas had been randomly interchanged. The subjects did better when the bicycle could be found in a natural context (see figure 16.5).

In order to explore the role of prior knowledge in the conjoining of properties. Deborah Butler and I did a further study of illusory conjunctions. We showed subjects a set of three colored objects flanked on each side by a digit. Then, some 200 milliseconds later, we showed them a pointer, which was accompanied by a random checkerboard in order to wipe out any visual persistence from the initial display. We asked the subjects to attend to the two digits and report them, and then to say which object the pointer had designated. The sequence was too brief to allow the subjects to focus their attention on all three objects.

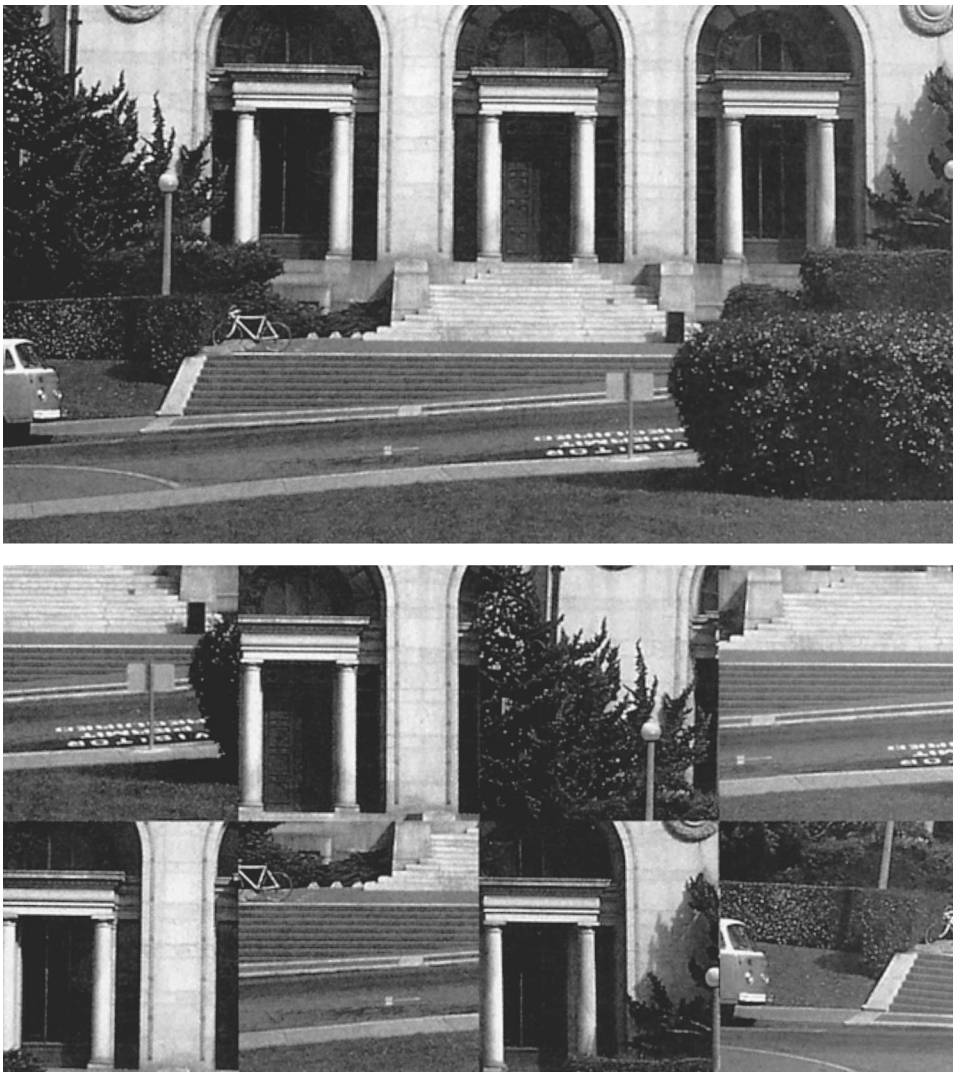


Figure 16.5
Prior knowledge as a guide in visual perception is tested by asking subjects to search for a familiar object in a photography of an unexceptional scene (top) and in a jumbled photograph of the scene (bottom). Here, the task is simply to find the bicycle. It tends to take longer in the jumbled image. The implication is that knowledge of the world (in this case, expectations about the characteristic locations of bicycles in urban landscapes) speeds up perception and makes it less subject to error. Certain early aspects of the information processing that underlies visual perception nonetheless seem to happen automatically: without the influence of prior knowledge. The illustration was modeled after experiments done by Irving Biederman of the State University of New York at Buffalo.

The crucial aspect of the experiment lay in the labels we gave the objects. We told one group of subjects that the display would consist of "an orange carrot, a blue lake, and a black tire." Occasional objects (one in four) were shown in the wrong color to ensure that the subjects could not just name the color they would know in advance ought to be associated with a given shape. For another group of subjects the same display was described as "an orange triangle, a blue ellipse, and a black ring."

The results were significant. The group given arbitrary pairings of colors and shapes reported many illusory conjunctions: 29 percent of their responses represented illusory recombinations of colors and shapes from the display, whereas 13 percent were reports of colors or shapes not present in the display. In contrast, the group expecting familiar objects saw rather few illusory conjunctions: They wrongly recombined colors and shapes only 5 percent more often than they reported colors and shapes not present in the display.

We occasionally gave a third group of subjects the wrong combinations when they were expecting most objects to be in their natural colors. To our surprise we found no evidence that subjects generated illusory conjunctions to fit their expectations. For example, they were no more likely to see the triangle (the "carrot") as orange when another object in the display was orange than they were when no orange was present. There seem to be two implications: Prior knowledge and expectations do indeed help one to use attention efficiently in conjoining features, but prior knowledge and expectations seem not to induce illusory exchanges of features to make abnormal objects normal again. Thus illusory conjunctions seem to arise at a stage of visual processing that precedes semantic access to knowledge of familiar objects. The conjunctions seem to be generated preattentively from the sensory data, bottom-up, and not to be influenced by top-down constraints.

How are objects perceived once attention has been focused on them and the correct set of properties has been selected from those present in the scene? In particular, how does one generate and maintain an object's perceptual unity even when objects move and change? Imagine a bird perched on a branch, seen from a particular angle and in a particular illumination. Now watch its shape, its size, and its color all change as it preens itself, opens its wings, and flies away. In spite of these major transformations in virtually all its properties, the bird retains its perceptual integrity: It remains the same single object.

Daniel Kahneman of the University of California at Berkeley and I have suggested that object perception is mediated not only by recognition, or matching to a stored label or description, but also by the construction of a temporary representation that is specific to the object's current appearance and is constantly updated as the object changes. We have drawn an analogy to a file in which all the perceptual information about a particular object is entered, just as the police might open a file on a particular crime, in which they collect all the information about the crime as the information accrues. The perceptual continuity of an object would then depend on its current manifestation being allocated to the same file as its earlier appearances. Such allocation is possible if the object remains stationary or if it changes location within constraints that allow the perceptual system to keep track of which file it should belong to.

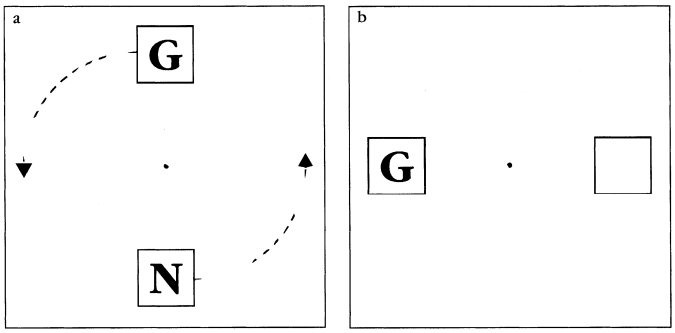


Figure 16.6
Integration of sensory information into what amounts to a file on each perceptual object was tested by the motion of frames. In each trial, two frames appeared, then two letters were briefly flashed in the frames (a). The frames moved to new locations, and a letter appeared in one of the two (b). The subject's task was to name the final letter as quickly as possible. If the final letter matched the initial letter and appeared in the same frame, the naming was faster than if the letter had appeared in the other frame or differed from the initial letter. The implication is that it takes more time to create or update a file on an object than it does simply to perceive the same object a second time.

In order to test this idea we joined with Brian Gibbs in devising a letter-naming task (figure 16.6). Two letters were briefly flashed in the centers of two frames. The empty frames then moved to new locations. Next, another letter appeared in one of the two frames. We devised the display so that the temporal and spatial separations between the priming letter and the final letter were always the same; the only thing that differed was the motion of the frames. The subjects' task was to name the final letter as quickly as possible.

We knew that the prior exposure to a given letter should normally lessen the time it takes to identify the same letter on a subsequent appearance; the effect is known as *priming*. The question that interested us was whether priming would occur only in particular circumstances. We argued that if the final letter is the same as the priming letter and appears in the same frame as the priming letter, the two should be seen as belonging to the same object; in this case, we could think of the perceptual task as simply re-viewing the original object in its shifted position. If, on the other hand, a new letter appears in the same frame, the object file should have to be updated, perhaps increasing the time it takes for subjects to become aware of the letter and name it.

Actually the priming was found to be object-specific: Subjects named the final letter some 30 milliseconds faster if the same letter had appeared previously in the same frame. They showed no such benefit if the same letter had appeared previously in the other frame. The result is consistent with the hypothesis that the later stages of visual perception integrate information from the early, feature-sensitive stages in temporary object-specific representations.

The overall scheme I propose for visual processing can be put in the form of a model (figure 16.7). The visual system begins by coding a certain number of simple and useful properties in what can be considered a stack of maps. In the brain such maps ordinarily preserve the spatial relations of the visual world itself. Nevertheless, the spatial information they contain may not be directly

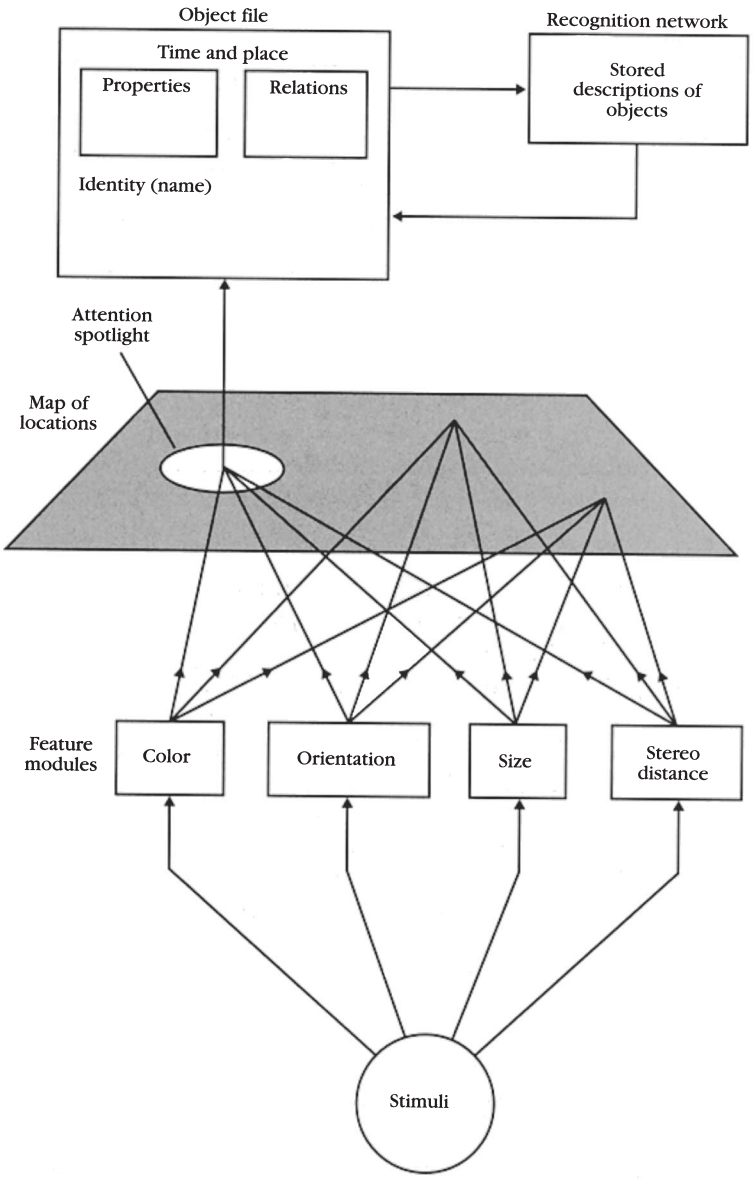


Figure 16.7
Hypothetical model of the early stages in visual perception emerges from the author's experiments. The model proposes that early vision encodes some simple and useful properties of a scene in a number of feature maps, which may preserve the spatial relations of the visual world but do not themselves make spatial information available to subsequent processing stages. Instead, focused attention (employing a master map of locations) selects and integrates the features present at particular locations. At later stages, the integrated information serves to create and update files on perceptual objects. In turn, the file contents are compared with descriptions stored in a recognition network. The network incorporates the attributes, behavior, names, and significance of familiar objects.

available to the subsequent stages of visual processing. Instead the presence of each feature may be signaled without a specification of *where* it is.

In the subsequent stages, focused attention acts. In particular, focused attention is taken to operate by means of a master map of locations, in which the presence of discontinuities in intensity or color is registered without specification of what the discontinuities are. Attention makes use of this master map, simultaneously selecting, by means of links to the separate feature maps, all the features that currently are present in a selected location. These are entered into a temporary object representation, or file.

Finally, the model posits that the integrated information about the properties and structural relations in each object file is compared with stored descriptions in a "recognition network." The network specifies the critical attributes of cats, trees, bacon and eggs, one's grandmothers, and all other familiar perceptual objects, allowing access to their names, their likely behavior, and their current significance. I assume that conscious awareness depends on the object files and on the information they contain. It depends, in other words, on representations that collect information about particular objects, both from the analyses of sensory features and from the recognition network, and continually update the information. If a significant discontinuity in space or time occurs, the original file on an object may be canceled: it ceases to be a source of perceptual experience. As for the object, it disappears and is replaced by a new object with its own new temporary file, ready to begin a new perceptual history.

PART IX

Human-Computer Interaction

Chapter 17

The Psychopathology of Everyday Things

Donald A. Norman

Kenneth Olsen, the engineer who founded and still runs Digital Equipment Corp., confessed at the annual meeting that he can't figure out how to heat a cup of coffee in the company's microwave oven.[1]

You Would Need an Engineering Degree to Figure This Out

"You would need an engineering degree from MIT to work this," someone once told me, shaking his head in puzzlement over his brand new digital watch. Well, I have an engineering degree from MIT. (Kenneth Olsen has two of them, and he can't figure out a microwave oven.) Give me a few hours and I can figure out the watch. But why should it take hours? I have talked with many people who can't use all the features of their washing machines or cameras, who can't figure out how to work a sewing machine or a video cassette recorder, who habitually turn on the wrong stove burner.

Why do we put up with the frustrations of everyday objects, with objects that we can't figure out how to use, with those neat plastic-wrapped packages that seem impossible to open, with doors that trap people, with washing machines and dryers that have become too confusing to use, with audio-stereo-television-video-cassette-recorders that claim in their advertisements to do everything, but that make it almost impossible to do anything?

The human mind is exquisitely tailored to make sense of the world. Give it the slightest clue and off it goes, providing explanation, rationalization, understanding. Consider the objects—books, radios, kitchen appliances, office machines, and light switches—that make up our everyday lives. Well-designed objects are easy to interpret and understand. They contain visible clues to their operation. Poorly designed objects (such as figure 17.1) can be difficult and frustrating to use. They provide no clues—or sometimes false clues. They trap the user and thwart the normal process of interpretation and understanding. Alas, poor design predominates. The result is a world filled with frustration, with objects that cannot be understood, with devices that lead to error. This chapter is an attempt to change things.

The Frustrations of Everyday Life

If I were placed in the cockpit of a modern jet airliner, my inability to perform gracefully and smoothly would neither surprise nor bother me. But I shouldn't

From chapter 1 in *The Design of Everyday Things* (New York: Doubleday, 1990), 1–34. Reprinted with permission.

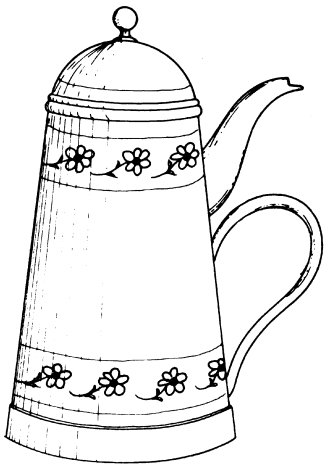


Figure 17.1
Carelman's Coffeepot for Masochists. The French artist Jacques Carelman in his series of books *Catalogue d'objets introuvables (Catalog of unfindable objects)* provides delightful examples of everyday things that are deliberately unworkable, outrageous, or otherwise ill-formed. Jacques Carelman: "Coffeepot for Masochists." Copyright © 1969–76–80 by Jacques Carelman and A. D. A. G. P. Paris. From Jacques Carelman, *Catalog of Unfindable Objects*, Balland, éditeur, Paris-France. Used by permission of the artist.

have trouble with doors and switches, water faucets and stoves. "Doors?" I can hear the reader saying, "you have trouble opening doors?" Yes. I push doors that are meant to be pulled, pull doors that should be pushed, and walk into doors that should be slid. Moreover, I see others having the same troubles—unnecessary troubles. There are psychological principles that can be followed to make these things understandable and usable.

Consider the door. There is not much you can do to a door: you can open it or shut it. Suppose you are in an office building, walking down a corridor. You come to a door. In which direction does it open? Should you pull or push, on the left or the right? Maybe the door slides. If so, in which direction? I have seen doors that slide up into the ceiling. A door poses only two essential questions: In which direction does it move? On which side should one work it? The answers should be given by the design, without any need for words or symbols, certainly without any need for trial and error.

A friend told me of the time he got trapped in the doorway of a post office in a European city. The entrance was an imposing row of perhaps six glass swinging doors, followed immediately by a second, identical row. That's a standard design: it helps reduce the airflow and thus maintain the indoor temperature of the building.

My friend pushed on the side of one of the leftmost pair of outer doors. It swung inward, and he entered the building. Then, before he could get to the next row of doors, he was distracted and turned around for an instant. He didn't realize it at the time, but he had moved slightly to the right. So when he came to the next door and pushed it, nothing happened. "Hmm," he thought, "must be locked." So he pushed the side of the adjacent door. Nothing. Puzzled, my friend decided to go outside again. He turned

around and pushed against the side of a door. Nothing. He pushed the adjacent door. Nothing. The door he had just entered no longer worked. He turned around once more and tried the inside doors again. Nothing. Concern, then mild panic. He was trapped! Just then, a group of people on the other side of the entranceway (to my friend's right) passed easily through both sets of doors. My friend hurried over to follow their path.

How could such a thing happen? A swinging door has two sides. One contains the supporting pillar and the hinge, the other is unsupported. To open the door, you must push on the unsupported edge. If you push on the hinge side, nothing happens. In this case, the designer aimed for beauty, not utility. No distracting lines, no visible pillars, no visible hinges. So how can the ordinary user know which side to push on? While distracted, my friend had moved toward the (invisible) supporting pillar, so he was pushing the doors on the hinged side. No wonder nothing happened. Pretty doors. Elegant. Probably won a design prize.

The door story illustrates one of the most important principles of design: *visibility*. The correct parts must be visible, and they must convey the correct message. With doors that push, the designer must provide signals that naturally indicate where to push. These need not destroy the aesthetics. Put a vertical plate on the side to be pushed, nothing on the other. Or make the supporting pillars visible. The vertical plate and supporting pillars are *natural* signals, *naturally* interpreted, without any need to be conscious of them. I call the use of natural signals *natural design* and elaborate on the approach throughout this chapter. Figure 17.2 illustrates a similar problem to the doors in the European post office. Go to "B".

Visibility problems come in many forms. My friend, trapped between the glass doors, suffered from a lack of clues that would indicate what part of a door should be operated. Other problems concern the *mappings* between what you want to do and what appears to be possible. Consider one type of slide projector. This projector has a single button to control whether the slide tray moves forward or backward. One button to do two things? What is the mapping? How can you figure out how to control the slides? You can't. Nothing is visible to give the slightest hint. Here is what happened to me in one of the many unfamiliar places I've lectured in during my travels as a professor:

The Leitz slide projector illustrated in figure 17.3 has shown up several times in my travels. The first time, it led to a rather dramatic incident. A conscientious student was in charge of showing my slides. I started my talk and showed the first slide. When I finished with the first slide and asked for the next, the student carefully pushed the control button and watched in dismay as the tray backed up, slid out of the projector and plopped off the table onto the floor, spilling its entire contents. We had to delay the lecture fifteen minutes while I struggled to reorganize the slides. It wasn't the student's fault. It was the fault of the elegant projector. With only one button to control the slide advance, how could one switch from forward to reverse? Neither of us could figure out how to make the control work.

All during the lecture the slides would sometimes go forward, sometimes backward. Afterward, we found the local technician, who explained it to us. A brief push of the button and the slide would go forward, a long push and it would reverse. (Pity the conscientious student who kept pushing it hard—and long—to make sure that the switch was making contact.) What an elegant design. Why, it managed to do two

Figure 17.2
A Row of Swinging Glass Doors in a Boston Hotel. A similar problem to the doors from that European post office. On which side of the door should you push? When I asked people who had just used the doors, most couldn't say. Yet only a few people I watched had trouble with the doors. The designers had incorporated a subtle clue into the design. Note that the horizontal bars are not centered: they are a bit closer together on the sides you should push on. The design almost works—but not entirely, for not everyone used the doors right on the first try.

Taste (7) für Diawechsel am Gerät

Diawechsel vorwärts = kurz drücken,

Diawechsel rückwärtz = länger drücken.

Button (7) for changing the slides

Slide change forward = short press,

Slide change backward = longer press.

Figure 17.3
Leitz Pravodit Slide Projector. I finally tracked down the instruction manual for that projector. A photograph of the projector has its parts numbered. The button for changing slides is number 7. The button itself has no labels. Who could discover this operation without the aid of the manual? Here is the entire text related to the button, in the original German and in my English translation.

functions with only one button! But how was a first-time user of the projector to know this?

As another example, consider the beautiful Amphithéâtre Louis-Laird in the Paris Sorbonne, which is filled with magnificent paintings of great figures in French intellectual history. (The mural on the ceiling shows lots of naked women floating about a man who is valiantly trying to read a book. The painting is right side up only for the lecturer—it is upside down for all the people in the audience.) The room is a delight to lecture in, at least until you ask for the projection screen to be lowered. "Ah," says the professor in charge, who gestures to the technician, who runs out of the room, up a short flight of stairs, and out of sight behind a solid wall. The screen comes down and stops. "No, no," shouts the professor, "a little bit more." The screen comes down again, this time too much. "No, no, no!" the professor jumps up and down and gestures wildly. It's a lovely room, with lovely paintings. But why can't the person who is trying to lower or raise the screen see what he is doing?

New telephone systems have proven to be another excellent example of incomprehensible design. No matter where I travel, I can count upon finding a particularly bad example.

When I visited Basic Books, I noticed a new telephone system. I asked people how they liked it. The question unleashed a torrent of abuse. "It doesn't have a hold function," one woman complained bitterly—the same complaint people at my university made about their rather different system. In older days, business phones always had a button labeled "hold." You could push the button and hang up the phone without losing the call on your line. Then you could talk to a colleague, or pick up another telephone call, or even pick up the call at another phone with the same telephone number. A light on the hold button indicated when the function was in use. It was an invaluable tool for business. Why didn't the new phones at Basic Books or in my university have a hold function, if it is so essential? Well, they did, even the very instrument the woman was complaining about. But there was no easy way to discover the fact, nor to learn how to use it.

I was visiting the University of Michigan and I asked about the new system there. "Yech!" was the response, "and it doesn't even have a hold function!" Here we go again. What is going on? The answer is simple: first, look at the instructions for hold. At the University of Michigan the phone company provided a little plate that fits over the keypad and reminds users of the functions and how to use them. I carefully unhooked one of the plates from the telephone and made a photocopy (figure 17.4). Can you understand how to use it? I can't. There is a "call hold" operation, but it doesn't make sense to me, not for the application that I just described.

The telephone hold situation illustrates a number of different problems. One of them is simply poor instructions, especially a failure to relate the new functions to the similarly named functions that people already know about. Second, and more serious, is the lack of visibility of the operation of the system. The new telephones, for all their added sophistication, lack both the hold button and the flashing light of the old ones. The hold is signified by an arbitrary action: dialing an arbitrary sequence of digits (*8, or *99, or what have you: it varies from one phone system to another). Third, there is no visible outcome of the operation.

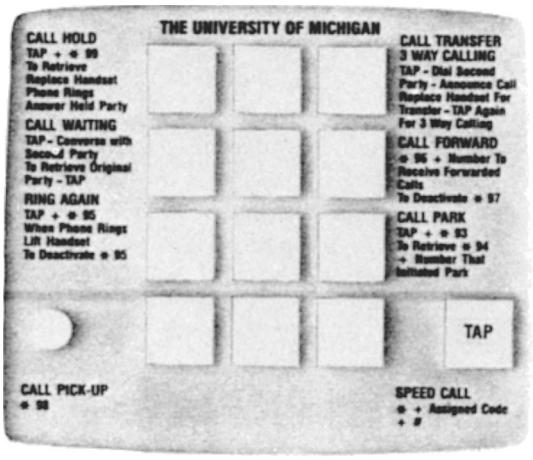


Figure 17.4
Plate Mounted over the Dial of the Telephones at the University of Michigan. These inadequate instructions are all that most users see. (The button labeled "TAP" at the lower right is used to transfer or pick up calls—it is pressed whenever the instruction plate says "TAP." The light on the lower left comes on whenever the telephone rings.)

Devices in the home have developed some related problems: functions and more functions, controls and more controls. I do not think that simple home appliances—stoves, washing machines, audio and television sets—should look like Hollywood's idea of a spaceship control room. They already do, much to the consternation of the consumer who, often as not, has lost (or cannot understand) the instruction manual, so—faced with the bewildering array of controls and displays—simply memorizes one or two fixed settings to approximate what is desired. The whole purpose of the design is lost.

In England I visited a home with a fancy new Italian washer-drier combination, with super-duper multi-symbol controls, all to do everything you ever wanted to do with the washing and drying of clothes. The husband (an engineering psychologist) said he refused to go near it. The wife (a physician) said she had simply memorized one setting and tried to ignore the rest.

Someone went to a lot of trouble to create that design. I read the instruction manual. That machine took into account everything about today's wide variety of synthetic and natural fabrics. The designers worked hard; they really cared. But obviously they had never thought of trying it out, or of watching anyone use it.

If the design was so bad, if the controls were so unusable, why did the couple purchase it? If people keep buying poorly designed products, manufacturers and designers will think they are doing the right thing and continue as usual.

The user needs help. Just the right things have to be visible: to indicate what parts operate and how, to indicate how the user is to interact with the device. Visibility indicates the mapping between intended actions and actual operations. Visibility indicates crucial distinctions—so that you can tell salt and pepper shakers apart, for example. And visibility of the effects of the operations tells you if the lights have turned on properly, if the projection screen has low-

ered to the correct height, or if the refrigerator temperature is adjusted correctly. It is lack of visibility that makes so many computer-controlled devices so difficult to operate. And it is an excess of visibility that makes the gadget-ridden, feature-laden modern audio set or video cassette recorder (VCR) so intimidating.

The Psychology of Everyday Things

This chapter is about the psychology of everyday things. POET emphasizes the understanding of everyday things, things with knobs and dials, controls and switches, lights and meters. The instances we have just examined demonstrate several principles, including the importance of visibility, appropriate clues, and feedback of one's actions. These principles constitute a form of psychology—the psychology of how people interact with things. A British designer once noted that the kinds of materials used in the construction of passenger shelters affected the way vandals responded. He suggested that there might be a psychology of materials.

Affordances

In one case, the reinforced glass used to panel shelters (for railroad passengers) erected by British Rail was smashed by vandals as fast as it was renewed. When the reinforced glass was replaced by plywood boarding, however, little further damage occurred, although no extra force would have been required to produce it. Thus British Rail managed to elevate the desire for defacement to those who could write, albeit in somewhat limited terms. Nobody has, as yet, considered whether there is a kind of psychology of materials. But on the evidence, there could well be![2]

There already exists the start of a psychology of materials and of things, the study of affordances of objects. When used in this sense, the term *affordance* refers to the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used (see figures 17.5 and 17.6). A chair affords ("is for") support and, therefore, affords sitting. A chair can also be carried. Glass is for seeing through, and for breaking. Wood is normally used for solidity, opacity, support, or carving. Flat, porous, smooth surfaces are for writing on. So wood is also for writing on. Hence the problem for British Rail: when the shelters had glass, vandals smashed it; when they had plywood, vandals wrote on and carved it. The planners were trapped by the affordances of their materials.[3]

Affordances provide strong clues to the operations of things. Plates are for pushing. Knobs are for turning. Slots are for inserting things into. Balls, are for throwing or bouncing. When affordances are taken advantage of, the user knows what to do just by looking: no picture, label, or instruction is required. Complex things may require explanation, but simple things should not. When simple things need pictures, labels, or instructions, the design has failed.

A psychology of causality is also at work as we use everyday things. Something that happens right after an action appears to be caused by that action. Touch a computer terminal just when it fails, and you are apt to believe that

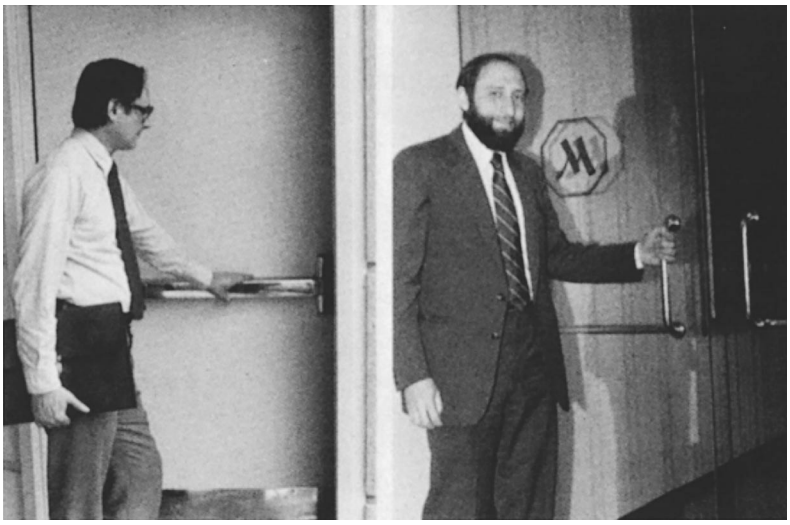


Figure 17.5
Affordances of Doors. Door hardware can signal whether to push or pull without signs. The flat horizontal bar of *A* (above left) affords no operations except pushing: it is excellent hardware for a door that must be pushed to be opened. The door in *B* (above right) has a different kind of bar on each side, one relatively small and vertical to signify a pull, the other relatively large and horizontal to signify a push. Both bars support the affordance of grasping: size and position specify whether the grasp is used to push or pull—though ambiguously.

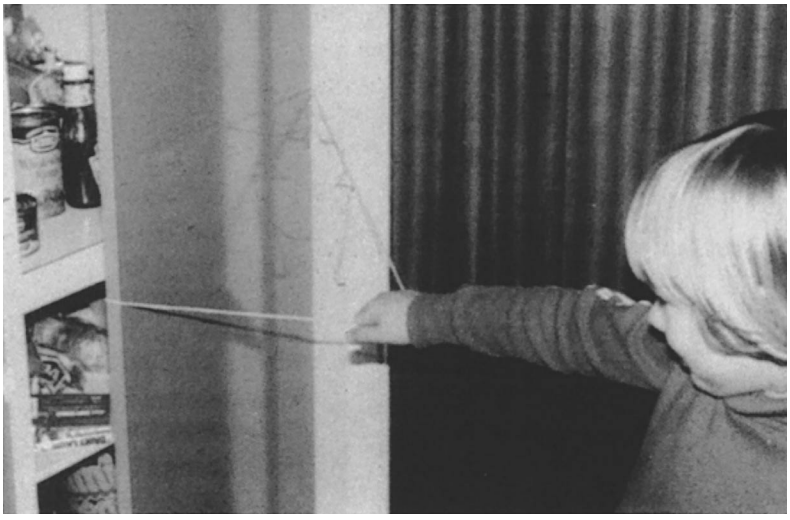


Figure 17.6
When Affordances Fail. I had to tie a string around my cabinet door to afford pulling.

you caused the failure, even though the failure and your action were related only by coincidence. Such false causality is the basis for much superstition. Many of the peculiar behaviors of people using computer systems or complex household appliances result from such false coincidences. When an action has no apparent result, you may conclude that the action was ineffective. So you repeat it. In earlier days, when computer word processors did not always show the results of their operations, people would sometimes attempt to change their manuscript, but the lack of visible effect from each action would make them think that their commands had not been executed, so they would repeat the commands, sometimes over and over, to their later astonishment and regret. It is a poor design that allows either kind of false causality to occur.

Twenty Thousand Everyday Things

There are an amazing number of everyday things, perhaps twenty thousand of them. Are there really that many? Start by looking about you. There are light fixtures, bulbs, and sockets; wall plates and screws; clocks, watches, and watchbands. There are writing devices (I count twelve in front of me, each different in function, color, or style). There are clothes, with different functions, openings, and flaps. Notice the variety of materials and pieces. Notice the variety of fasteners—buttons, zippers, snaps, laces. Look at all the furniture and food utensils: all those details, each serving some function for manufacturability, usage, or appearance. Consider the work area: paper clips, scissors, pads of paper, magazines, books, bookmarks. In the room I'm working in, I counted more than a hundred specialized objects before I tired. Each is simple, but each requires its own method of operation, each has to be learned, each does its own specialized task, and each has to be designed separately. Furthermore, many of the objects are made of many parts. A desk stapler has sixteen parts, a household iron fifteen, the simple bathtub-shower combination twenty-three. You can't believe these simple objects have so many parts? Here are the eleven basic parts to a sink: drain, flange (around the drain), pop-up stopper, basin, soap dish, overflow vent, spout, lift rod, fittings, hot-water handle, and cold-water handle. We can count even more if we start taking the faucets, fittings, and lift rods apart.

The book *What's What: A Visual Glossary of the Physical World* has more than fifteen hundred drawings and pictures and illustrates twenty-three thousand items or parts of items.[4] Irving Biederman, a psychologist who studies visual perception, estimates that there are probably "30,000 readily discriminable objects for the adult."[5] Whatever the exact number, it is clear that the difficulties of everyday life are amplified by the sheer profusion of items. Suppose that each everyday thing takes only one minute to learn; learning 20,000 of them occupies 20,000 minutes—333 hours or about 8 forty-hour work weeks. Furthermore, we often encounter new objects unexpectedly, when we are really concerned with something else. We are confused and distracted and what ought to be a simple, effortless, everyday thing interferes with the important task of the moment.

How do people cope? Part of the answer lies in the way the mind works—in the psychology of human thought and cognition. Part lies in the information available from the appearance of the objects—the psychology of everyday

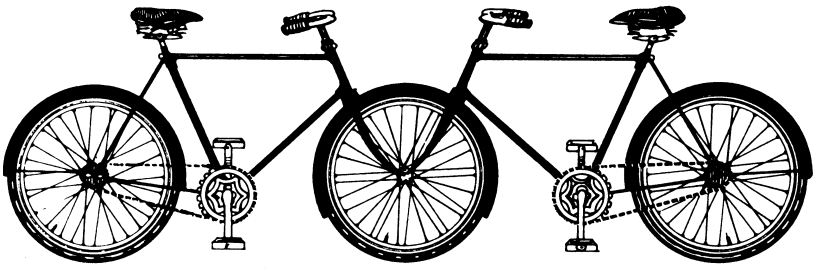


Figure 17.7
Carelman's Tandem "Convergent Bicycle (Model for Fiancés)." Jacques Carelman: "Convergent Bicycle" Copyright © 1969–76–80 by Jacques Carelman and A. D. A. G. P. Paris. From Jacques Carelman, *Catalog of Unfindable Objects*, Balland, éditeur, Paris-France. Used by permission of the artist.

things. And part comes from the ability of the designer to make the operation clear, to project a good image of the operation, and to take advantage of other things people might be expected to know. Here is where the designer's knowledge of the psychology of people coupled with knowledge of how things work becomes crucial.

Conceptual Models

Consider the rather strange bicycle illustrated in figure 17.7. You know it won't work because you form a *conceptual model* of the device and mentally simulate its operation. You can do the simulation because the parts are visible and the implications clear.

Other clues to how things work come from their visible structure—in particular from *affordances, constraints*, and *mappings*. Consider a pair of scissors: even if you have never seen or used them before, you can see that the number of possible actions is limited. The holes are clearly there to put something into, and the only logical things that will fit are fingers. The holes are affordances: they allow the the fingers to be inserted. The sizes of the holes provide *constraints* to limit the possible fingers: the big hole suggests several fingers, the small hole only one. The mapping between holes and fingers—the set of possible operations—is suggested and constrained by the holes. Moreover, the operation is not sensitive to finger placement: if you use the wrong fingers, the scissors still work. You can figure out the scissors because their operating parts are visible and the implications clear. The conceptual model is made obvious, and there is effective use of affordances and constraints.

As a counterexample, consider the digital watch, one with two to four push buttons on the front or side. What are those push buttons for? How would you set the time? There is no way to tell—no evident relationship between the operating controls and the functions, no constraints, no apparent mappings. With the scissors, moving the handle makes the blades move. The watch and the Leitz slide projector provide no visible relationship between the buttons and the possible actions, no discernible relationship between the actions and the end result.

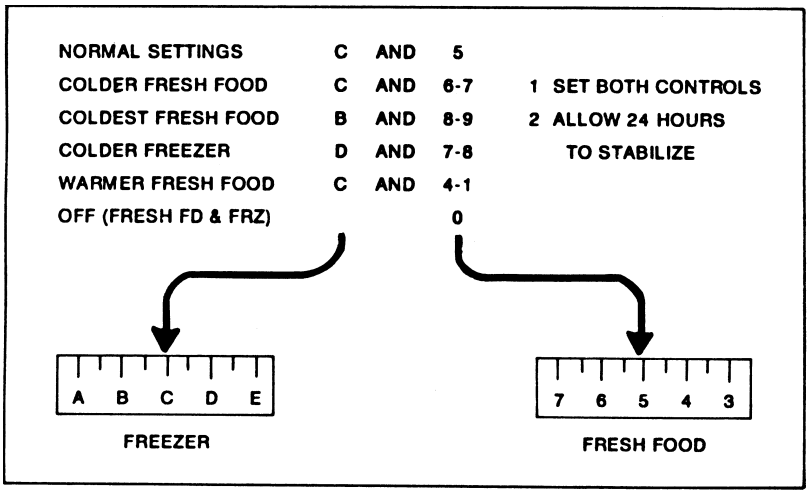


Figure 17.8
My Refrigerator. Two compartments—fresh food and freezer—and two controls (in the fresh food unit). The illustration shows the controls and instructions. Your task: Suppose the freezer is too cold, the fresh food section is just right. How would you adjust the controls so as to make the freezer warmer and keep the fresh food the same? (From Norman, 1986.)

Principles of Design for Understandability and Usability

We have now encountered the fundamental principles of designing for people: (1) provide a good conceptual model and (2) make things visible.

Provide a Good Conceptual Model

A good conceptual model allows us to predict the effects of our actions. Without a good model we operate by rote, blindly; we do operations as we were told to do them; we can't fully appreciate why, what effects to expect, or what to do if things go wrong. As long as things work properly, we can manage. When things go wrong, however, or when we come upon a novel situation, then we need a deeper understanding, a good model.

For everyday things, conceptual models need not be very complex. After all, scissors, pens, and light switches are pretty simple devices. There is no need to understand the underlying physics or chemistry of each device we own, simply the relationship between the controls and the outcomes. When the model presented to us is inadequate or wrong (or, worse, nonexistent), we can have difficulties. Let me tell you about my refrigerator.

My house has an ordinary, two-compartment refrigerator—nothing very fancy about it. The problem is that I can't set the temperature properly. There are only two things to do: adjust the temperature of the freezer compartment and adjust the temperature of the fresh food compartment. And there are two controls, one labeled "freezer," the other "fresh food." What's the problem?

You try it. Figure 17.8 shows the instruction plate from inside the refrigerator. Now, suppose the freezer is too cold, the fresh food section just right. You want to make

the freezer warmer, keeping the fresh food constant. Go on, read the instructions, figure them out.

Oh, perhaps I'd better warn you. The two controls are not independent. The freezer control affects the fresh food temperature, and the fresh food control affects the freezer. And don't forget to wait twenty-four hours to check on whether you made the right adjustment, if you can remember what you did.

Control of the refrigerator is made difficult because the manufacturer provides a false conceptual model. There are two compartments and two controls. The setup clearly and unambiguously provides a simple model for the user: each control is responsible for the temperature of the compartment that carries its name. Wrong. In fact, there is only one thermostat and only one cooling mechanism. One control adjusts the thermostat setting, the other the relative proportion of cold air sent to each of the two compartments of the refrigerator. This is why the two controls interact. With the conceptual model provided by the manufacturer, adjusting the temperatures is almost impossible and always frustrating. Given the correct model, life would be much easier (figure 17.9).

Why did the manufacturer present the wrong conceptual model? Perhaps the designers thought the correct model was too complex, that the model they were giving was easier to understand. But with the wrong conceptual model, it is impossible to set the controls. And even though I am convinced I now know the correct model, I still cannot accurately adjust the temperatures because the refrigerator design makes it impossible for me to discover which control is for the thermostat, which control is for the relative proportion of cold air, and in which compartment the thermostat is located. The lack of immediate feedback for the actions does not help: with a delay of twenty-four hours, who can remember what was tried?

The topic of conceptual models will reappear in the book. They are part of an important concept in design: *mental models*, the models people have of themselves, others, the environment, and the things with which they interact. People form mental models through experience, training, and instruction. The mental model of a device is formed largely by interpreting its perceived actions and its visible structure. I call the visible part of the device the *system image* (figure 17.10). When the system image is incoherent or inappropriate, as in the case of the refrigerator, then the user cannot easily use the device. If it is incomplate or contradictory, there will be trouble.

Make Things Visible

The problems caused by inadequate attention to visibility are all neatly demonstrated with one simple appliance: the modern telephone.

I stand at the blackboard in my office, talking with a student, when my telephone rings. Once, twice it rings. I pause, trying to complete my sentence before answering. The ringing stops. "I'm sorry," says the student. "Not your fault," I say. "But it's no problem, the call now transfers to my secretary's phone. She'll answer it." As we listen we hear her phone start to ring. Once, twice. I look at my watch. Six o'clock: it's late, the office staff has left for the day. I rush out of my office to my secretary's phone, but as I get there, it stops ringing. "Ah," I think, "it's being transferred to another phone." Sure enough, the phone in the adjacent office now starts ringing. I rush to that office,

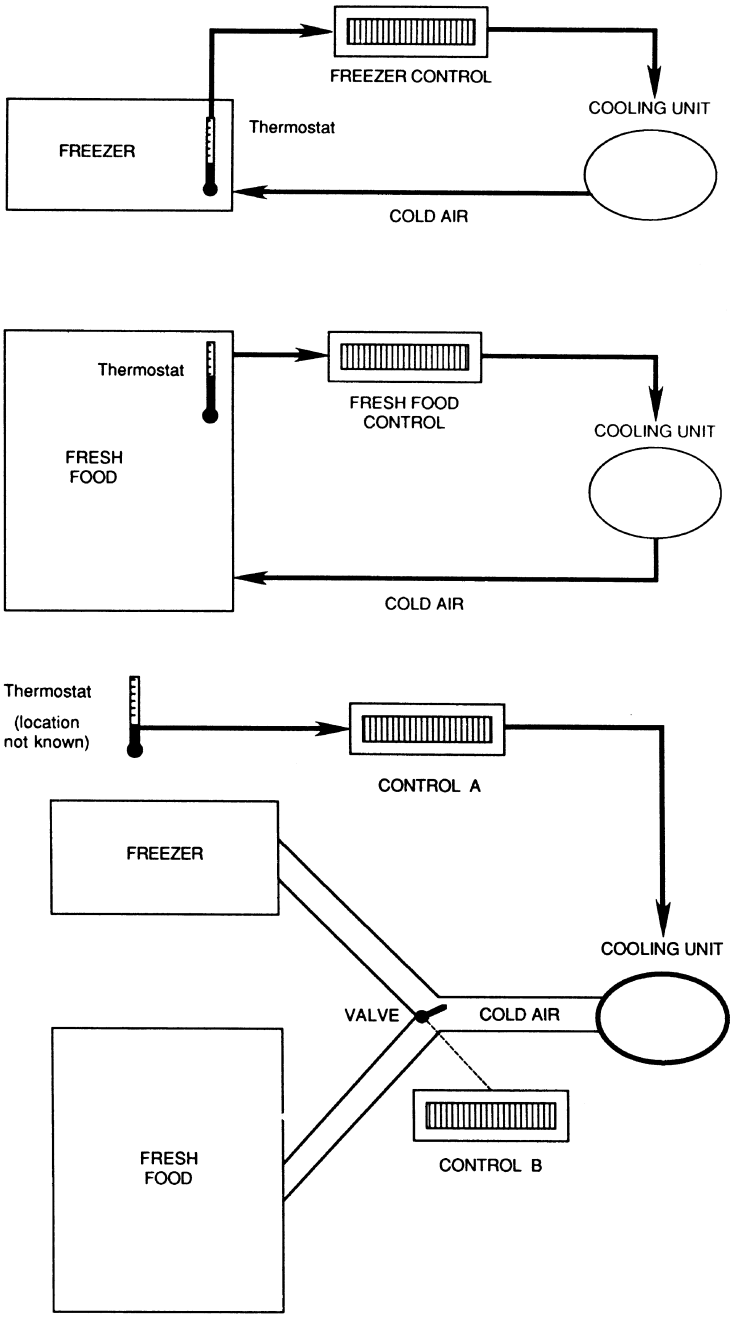


Figure 17.9
Two Conceptual Models for My Refrigerator. The model *A* (above) is provided by the system image on the refrigerator as gleaned from the controls and instructions; *B* (below) is the correct conceptual model. The problem is that it is impossible to tell in which compartment the thermostat is located and whether the two controls are in the freezer and fresh food compartment, or vice versa.

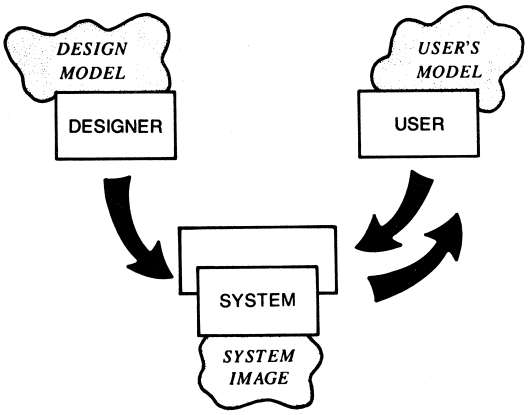


Figure 17.10
Conceptual Models. The *design model* is the designer's conceptual method. The *user's model* is the mental model developed through interaction with the system. The *system image* results from the physical structure that has been built (including documentation, instructions, and labels). The designer expects the user's model to be identical to the design model. But the designer doesn't talk directly with the user—all communication takes place through the system image. If the system image does not make the design model clear and consistent, then the user will end up with the wrong mental model. (From Norman, 1986.)

but it is locked. Back to my office to get the key, out to the locked door, fumble with the lock, into the office, and to the now quiet phone. I hear a telephone down the hall start to ring. Could that still be my call, making its way mysteriously, with a predetermined lurching path, through the phones of the building? Or is it just another telephone call coincidentally arriving at this time?

In fact, I could have retrieved the call from my office, had I acted quickly enough. The manual states: "Within your pre-programmed pick-up group, dial 14 to connect to incoming call. Otherwise, to answer any ringing extension, dial ringing extension number, listen for busy tone. Dial 8 to connect to incoming call." Huh? What do those instructions mean? What is a "pre-programmed pick-up group," and why do I even want to know? What is the extension number of the ringing phone? Can I remember all those instructions when I need them? No.

Telephone chase is the new game in the modem office, as the automatic features of telephones go awry—features designed without proper thought, and certainly without testing them with their intended users. There are several other games, too. One game is announced by the plea, "How do I answer this call?" The question is properly whined in front of a ringing, flashing telephone, receiver in hand. Then there is the paradoxical game entitled "This telephone doesn't have a hold function." The accusation is directed at a telephone that actually *does* have a hold function. And, finally, there is "What do you mean I called you, you called me!"

Many of the modern telephone systems have a new feature that automatically keeps trying to dial a number for you. This feature resides under names such as automatic

redialing or automatic callback. I am supposed to use this feature whenever I call someone who doesn't answer or whose line is busy. When the person next hangs up the phone, my phone will dial it again. Several automatic callbacks can be active at a time. Here's how it works. I place a phone call. There's no answer, so I activate the automatic callback feature. Several hours later my telephone rings. I pick it up and say "Hello," only to hear a ringing sound and then someone else saying "Hello."

"Hello," I answer, "who is this?"

"Who is this?" I hear in reply, "you called me."

"No," I say, "you called me, my phone just rang."

Slowly I realize that perhaps this is my delayed call. Now, let me see, who was I trying to call several hours ago? Did I have several callbacks in place? Why was I making the call?

The modern telephone did not happen by accident: it was carefully designed. Someone—more likely a team of people—invented a list of features thought desirable, invented what seemed to them to be plausible ways of controlling the features, and then put it all together. My university, focusing on cost and perhaps dazzled by the features, bought the system, spending millions of dollars on a telephone installation that has proved vastly unpopular and even unworkable. Why did the university buy the system? The purchase took several years of committee work and studies and presentations by competing telephone companies, and piles of documentation and specification. I myself took part, looking at the interaction between the telephone system and the computer networks, ensuring that the two would be compatible and reasonable in price. To my knowledge, nobody ever thought of trying out the telephones in advance. Nobody suggested installing them in a sample office to see whether users' needs would be met or whether users could understand how to operate the phone. The result: disaster. The main culprit—lack of visibility—was coupled with a secondary culprit—a poor conceptual model. Any money saved on the installation and purchase is quickly disappearing in training costs, missed calls, and frustration. Yet from what I have seen, the competing phone systems would not have been any better.

I recently spent six months at the Applied Psychology Unit in Cambridge, England. Just before I arrived the British Telecom Company had installed a new telephone system. It had lots and lots of features. The telephone instrument itself was unremarkable (figure 17.11). It was the standard twelve-button, push-button phone, except that it had an extra key labeled "R" off on the side. (I never did find out what that key did.)

The telephone system was a standing joke. Nobody could use all the features. One person even started a small research project to record people's confusions. Another person wrote a small "expert systems" computer program, one of the new toys of the field of artificial intelligence; the program can reason through complex situations. If you wanted to use the phone system, perhaps to make a conference call among three people, you asked the expert system and it would explain how to do it. So, you're on the line with someone and you need to add a third person to the call. First turn on your computer. Then load the expert system. After three or four minutes (needed for loading the program), type in what you want to accomplish. Eventually the computer will tell you what

Figure 17.11
British Telecom Telephone. This was in my office at the Applied Psychology Unit in Cambridge, England. It certainly looks simple, doesn't it?

to do—if you can remember why you want to do it, and if the person on the other end of the line is still around. But, as it happens, using the expert system is a lot easier than reading and understanding the manual provided with the telephone (figure 17.12).

Why is that telephone system so hard to understand? Nothing in it is conceptually difficult. Each of the operations is actually quite simple. A few digits to dial, that's all. The telephone doesn't even look complicated. There are only fifteen controls: the usual twelve buttons—ten labeled 0 through 9, #, and *—plus the handset itself, the handset button, and the mysterious "R" button. All except the "R" are the everyday parts of a normal modern telephone. Why was the system so difficult?

A designer who works for a telephone company told me the following story:

"I was involved in designing the faceplate of some of those new multifunction phones, some of which have buttons labeled 'R.' The 'R' button is kind of a vestigial feature. It is very hard to remove features of a newly designed product that had existed in an earlier version. It's kind of like physical evolution. If a feature is in the genome, and if that feature is not associated with any negativity (i.e., no customers gripe about it), then the feature hangs on for generations.

"It is interesting that things like the 'R' button are largely determined through examples. Somebody asks, 'What is the "R" button used for?' and the answer is to give an example: 'You can push "R" to access loudspeaker paging.' If nobody can think of an example, the feature is dropped. Designers are pretty bright people, however. They can come up with a plausible-sounding example for almost anything. Hence, you get features, many many features, and these features hang on for a long time. The end result is complex interfaces for essentially simple things."[6]

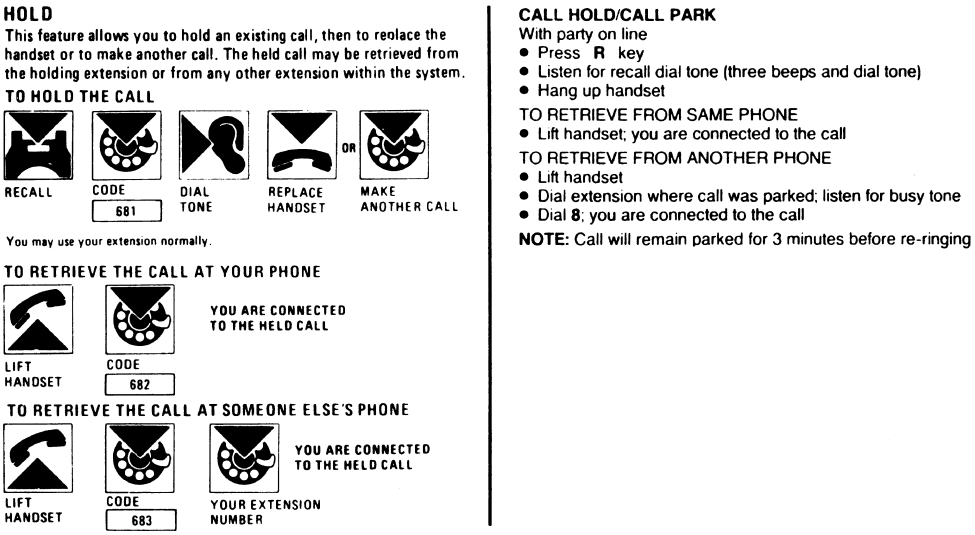


Figure 17.12
Two Ways to Use Hold on Modern Telephones. Illustration *A* (left) is the instruction manual page for British Telecom. The procedure seems especially complicated, with three 3-digit codes to be learned: 681, 682, and 683. Illustration *B* (right) shows the equivalent instructions for the Ericsson Single Line Analog Telephone installed at the University of California, San Diego. I find the second set of instructions easier to understand, but one must still dial an arbitrary digit: 8 in this case.

As I pondered this problem, I decided it would make sense to compare the phone system with something that was of equal or greater complexity but easier to use. So let us temporarily leave the difficult telephone system and take a look at my automobile. I bought a car in Europe. When I picked up the new car at the factory, a man from the company sat in the car with me and went over each control, explaining its function. When he had gone through the controls once, I said fine, thanked him, and drove away. That was all the instruction it took. There are 112 controls inside the car. This isn't quite as bad as it sounds. Twenty-five of them are on the radio. Another 7 are the temperature control system, and 11 work the windows and sunroof. The trip computer has 14 buttons, each matched with a specific function. So four devices—the radio, temperature controls, windows, and trip computer—have together 57 controls, or just over 50 percent of the ones available.

Why is the automobile, with all its varied functions and numerous controls, so much easier to learn and to use than the telephone system, with its much smaller set of functions and controls? What is good about the design of the car? Things are visible. There are good mappings, natural relationships, between the controls and the things controlled. Single controls often have single functions. There is good feedback. The system is understandable. In general, the relationships among the user's intentions, the required actions, and the results are sensible, nonarbitrary, and meaningful.

What is bad about the design of the telephone? There is no visible structure. Mappings are arbitrary: there is no rhyme or reason to the relationship between

the actions the user must perform and the results to be accomplished. The controls have multiple functions. There isn't good feedback, so the user is never sure whether the desired result has been obtained. The system, in general, is not understandable; its capabilities aren't apparent. In general, the relationships among the user's intentions, the required actions, and the results are completely arbitrary.

Whenever the number of possible actions exceeds the number of controls, there is apt to be difficulty. The telephone system has twenty-four functions, yet only fifteen controls—none of them labeled for specific action. In contrast, the trip computer for the car performs seventeen functions with fourteen controls. With minor exceptions, there is one control for each function. In fact, the controls with more than one function are indeed harder to remember and use. When the number of controls equals the number of functions, each control can be specialized, each can be labeled. The possible functions are visible, for each corresponds with a control. If the user forgets the functions, the controls serve as reminders. When, as on the telephone, there are more functions than controls, labeling becomes difficult or impossible. There is nothing to remind the user. Functions are invisible, hidden from sight. No wonder the operation becomes mysterious and difficult. The controls for the car are visible and, through their location and mode of operation, bear an intelligent relationship to their action. Visibility acts as a good reminder of what can be done and allows the control to specify how the action is to be performed. The good relationship between the placement of the control and what it does makes it easy to find the appropriate control for a task. As a result, there is little to remember.

The Principle of Mapping

Mapping is a technical term meaning the relationship between two things, in this case between the controls and their movements and the results in the world. Consider the mapping relationships involved in steering a car. To turn the car to the right, one turns the steering wheel clockwise (so that its top moves to the right). The user must identify two mappings here: one of the 112 controls affects the steering, and the steering wheel must be turned in one of two directions. Both are somewhat arbitrary. But the wheel and the clockwise direction are natural choices: visible, closely related to the desired outcome, and providing immediate feedback. The mapping is easily learned and always remembered.

Natural mapping, by which I mean taking advantage of physical analogies and cultural standards, leads to immediate understanding. For example, a designer can use spatial analogy: to move an object up, move the control up. To control an array of lights, arrange the controls in the same pattern as the lights. Some natural mappings are cultural or biological, as in the universal standard that a rising level represents more, a diminishing level, less. Similarly, a louder sound can mean a greater amount. Amount and loudness (and weight, line length, and brightness) are additive dimensions: add more to show incremental increases. Note that the logically plausible relationship between musical pitch and amount does not work: Would a higher pitch mean less or more of something? Pitch (and taste, color, and location) are substitutive dimensions: substitute one value for another to make a change. There is no natural concept of

Figure 17.13
Sear Adjustment Control from a Mercedes-Benz Automobile. This is an excellent example of natural mapping. The control is in the shape of the seat itself: the mapping is straightforward. To move the front edge of the seat higher, lift up on the front part of the button. To make the seat back recline, move the button back. Mercedes-Benz automobiles are obviously not everyday things for most people, but the principle doesn't require great expense or wealth. The same principle could be applied to much more common objects.

more or less in the comparison of different pitches, or hues, or taste qualities. Other natural mappings follow from the principles of perception and allow for the natural grouping or patterning of controls and feedback (see figure 17.13).

Mapping problems are abundant, one of the fundamental causes of difficulties. Consider the telephone. Suppose you wish to activate the callback on "no reply" function. To initiate this feature on one telephone system, press and release the "recall" button (the button on the handset), then dial 60, then dial the number you called.

There are several problems here. First, the description of the function is relatively complex—yet incomplete: What if two people set up callback at the same time? What if the person does not come back until a week later? What if you have meanwhile set up three or four other functions? What if you want to cancel it? Second, the action to be performed is arbitrary. (Dial 60. Why 60? Why not 73 or 27? How does one remember an arbitrary number?) Third, the sequence ends with what appears to be a redundant, unnecessary action: dialing the number of the person to be called. If the phone system is smart enough to do all these other things, why can't it remember the number that was just attempted; why must it be told all over again? And finally, consider the lack of feedback. How do I know I did the right action? Maybe I disconnected the phone. Maybe I set up some other special feature. There is no visible or audible way to know immediately.

A device is easy to use when there is visibility to the set of possible actions, where the controls and displays exploit natural mappings. The principles are simple but rarely incorporated into design. Good design takes care, planning, thought. It takes conscious attention to the needs of the user. And sometimes the designer gets it right:

Once, when I was at a conference at Gmunden, Austria, a group of us went off to see the sights. I sat directly behind the driver of the brand new, sleek, high-technology German tour bus. I gazed in wonder at the hundreds of controls scattered all over the front of the bus.

"How can you ever learn all those controls?" I asked the driver (with the aid of a German-speaking colleague). The driver was clearly puzzled by the question.

"What do you mean?" he replied. "Each control is just where it ought to be. There is no difficulty."

A good principle, that. Controls are where they ought to be. One function, one control. Harder to do, of course, than to say, but essentially this is the principle of natural mappings: the relationship between controls and actions should be apparent to the user. The problem of determining the "naturalness" of mappings is difficult, but crucial.

I've already described how my car's controls are generally easy to use. Actually, the car has lots of problems. The approach to usability used in the car seems to be to make sure that you can reach everything and see everything. That's good, but not nearly good enough.

Here is a simple example: the controls for the loudspeakers—a simple control that determines whether the sound comes out of the front speakers, the rear, or a combination (figure 17.14). Rotate the wheel from left to right or right to left. Simple, except how do you know which way to rotate the control? Which direction moves the sound to the rear, which to the front? If you want sound to come out of the front speaker, you should be able to move the control to the front. To get it out of the back, move the control to the back. Then the form of the motion would mimic the function and make a natural mapping. But the way the control is actually mounted in the car, forward and backward get translated into left and right. Which direction is which? There is no natural relationship. What's worse, the control isn't even labeled. Even the instruction manual does not say how to use it.

The control should be mounted so that it moves forward and backward. If that can't be done, rotate the control 90° on the panel so that it moves vertically. Moving something up to represent forward is not as natural as moving it forward, but at least it follows a standard convention.

In fact, we see that both the car and the telephone have easy functions and difficult ones. The car seems to have more of the easy ones, the telephone more of the difficult ones. Moreover, with the car, enough of the controls are easy that I can do almost everything I need to. Not so with the telephone: it is very difficult to use even a single one of the special features.

The easy things on both telephone and car have a lot in common, as do the difficult things. When things are visible, they tend to be easier than when they are not. In addition, there must be a close, *natural* relationship between the control and its function: *a natural mapping*.

Figure 17.14
The Front/Rear Speaker Selector of an Automobile Radio. Rotating the knob with the pictures of the speaker at either side makes the sound come entirely out of the front speakers (when the knob is all the way over to one side), entirely out of the rear speakers (when the knob is all the way the other way), or equally out of both (when the knob is midway). Which way is front, which rear? You can't tell by looking. While you're at it, imagine trying to manipulate the radio controls while keeping your eyes on the road.

The Principle of Feedback

Feedback—sending back to the user information about what action has actually been done, what result has been accomplished—is a well-known concept in the science of control and information theory. Imagine trying to talk to someone when you cannot even hear your own voice, or trying to draw a picture with a pencil that leaves no mark: there would be no feedback.

In the good old days of the telephone, before the American telephone system was divided among competing companies, before telephones were fancy and had so many features, telephones were designed with much more care and concern for the user. Designers at the Bell Telephone Laboratories worried a lot about feedback. The push buttons were designed to give an appropriate feel—tactile feedback. When a button was pushed, a tone was fed back into the earpiece so the user could tell that the button had been properly pushed. When the phone call was being connected, clicks, tones, and other noises gave the user feedback about the progress of the call. And the speaker's voice was always fed back to the earpiece in a carefully controlled amount, because the auditory feedback (called "sidetone") helped the person regulate how loudly to talk. All this has changed. We now have telephones that are much more powerful and often cheaper than those that existed just a few years ago—more function for less money. To be fair, these new designs are pushing hard on the paradox of technology: added functionality generally comes along at the price of added complexity. But that does not justify backward progress.

Why are the modern telephone systems so difficult to learn and to use? Basically, the problem is that the systems have more features and less feedback. Suppose all telephones had a small display screen, not unlike the ones on small, inexpensive calculators. The display could be used to present, upon the push of a button, a brief menu of all the features of the telephone, one by one. When the desired one was encountered, the user would push another button to indicate that it should be invoked. If further action was required, the display could tell the person what to do. The display could even be auditory, with speech instead of a visual display. Only two buttons need be added to the telephone: one to change the display, one to accept the option on display. Of course, the telephone would be slightly more expensive. The tradeoff is cost versus usability.[7]

Pity the Poor Designer

Designing well is not easy. The manufacturer wants something that can be produced economically. The store wants something that will be attractive to its customers. The purchaser has several demands. In the store, the purchaser focuses on price and appearance, and perhaps on prestige value. At home, the same person will pay more attention to functionality and usability. The repair service cares about maintainability: how easy is the device to take apart, diagnose, and service? The needs of those concerned are different and often conflict. Nonetheless, the designer may be able to satisfy everyone.

A simple example of good design is the $3\frac{1}{2}$-inch magnetic diskette for computers, a small circle of "floppy" magnetic material encased in hard plastic. Earlier types of floppy disks did not have this plastic case, which protects the magnetic material from abuse and damage. A sliding metal cover protects the delicate magnetic surface when the diskette is not in use and automatically opens when the diskette is inserted into the computer. The diskette has a square shape: there are apparently eight possible ways to insert it into the machine, only one of which is correct. What happens if I do it wrong? I try inserting the disk sideways. Ah, the designer thought of that. A little study shows that the case really isn't square: it's rectangular, so you can't insert a longer side. I try backward. The diskette goes in only part of the way. Small protrusions, indentations, and cutouts prevent the diskette from being inserted backward or upside down: of the eight ways one might try to insert the diskette, only one is correct, and only that one will fit. An excellent design.

Take another example of good design. My felt-tipped marking pen has ribs along only one of its sides; otherwise all sides look identical. Careful examination shows that the tip of the marker is angled and makes the best line if the marker is held with the ribbed side up, a natural result if the forefinger rests upon the ribs. No harm results if I hold the marker another way, but the marker writes less well. The ribs are a subtle design cue—functional, yet visibly and aesthetically unobtrusive.

The world is permeated with small examples of good design, with the amazing details that make important differences in our lives. Each detail was added by some person, a designer, carefully thinking through the uses of the device, the ways that people abuse things, the kinds of errors that can get made, and the functions that people wish to have performed.

Then why is it that so many good design ideas don't find their way into products in the marketplace? Or something good shows up for a short time, only to fall into oblivion? I once spoke with a designer about the frustrations of trying to get the best product out:

It usually takes five or six attempts to get a product right. This may be acceptable in an established product, but consider what it means in a new one. Suppose a company wants to make a product that will perhaps make a real difference. The problem is that if the product is truly revolutionary, it is unlikely that anyone will quite know how to design it right the first time; it will take several tries. But if a product is introduced into the marketplace and fails, well that is it. Perhaps it could be introduced a second time, or maybe even a third time, but after that it is dead: everyone believes it to be a failure.

I asked him to explain. "You mean," I said, "that it takes five or six tries to get an idea right?"

"Yes," he said, "at least that."

"But," I replied, "you also said that if a newly introduced product doesn't catch on in the first two or three times, then it is dead?"

"Yup," he said.

"Then new products are almost guaranteed to fail, no matter how good the idea."

"Now you understand," said the designer. "Consider the use of voice messages on complex devices such as cameras, soft-drink machines, and copiers. A failure. No longer even tried. Too bad. It really is a good idea, for it can be very useful when the hands or eyes are busy elsewhere. But those first few attempts were very badly done and the public scoffed—properly. Now, nobody dares try it again, even in those places where it is needed."

The Paradox of Technology

Technology offers the potential to make life easier and more enjoyable; each new technology provides increased benefits. At the same time, added complexities arise to increase our difficulty and frustration. The development of a technology tends to follow a U-shaped curve of complexity: starting high; dropping to a low, comfortable level; then climbing again. New kinds of devices are complex and difficult to use. As technicians become more competent and an industry matures, devices become simpler, more reliable, and more powerful. But then, after the industry has stabilized, newcomers figure out how to add increased power and capability, but always at the expense of added complexity and sometimes decreased reliability. We can see the curve of complexity in the history of the watch, radio, telephone, and television set. Take the radio. In the early days, radios were quite complex. To tune in a station required several adjustments, including one for the antenna, one for the radio frequency, one for intermediate frequencies, and controls for both sensitivity and loudness. Later radios were simpler and had controls only to turn it on, tune the station, and adjust the loudness. But the latest radios are again very complex, perhaps even more so than early ones. Now the radio is called a tuner, and it is littered with numerous controls, switches, slide bars, lights, displays, and meters. The modern sets are technologically superior, offering

higher quality sound, better reception, and enhanced capability. But what good is the technology if it is too complex to use?

The design problem posed by technological advances is enormous. Consider the watch. A few decades ago, watches were simple. All you had to do was set the time and keep them wound. The standard control was the stem: a knob at the side of the watch. Turning the knob wound the spring that worked the watch. Pulling the knob out and turning it made the hands move. The operations were easy to learn and easy to do. There was a reasonable relation between the turning of the knob and the resulting turning of the hands. The design even took into account human error: the normal position of the stem was for winding the spring, so that an accidental turn would not reset the time.

In the modern digital watch the spring is gone, replaced by a motor run by long-lasting batteries. All that remains is the task of setting the watch. The stem is still a sensible solution, for you can go fast or slow, forward or backward, until the exact desired time is reached. But the stem is more complex (and therefore more expensive) than simple push-button switches. If the only change in the transition from the spring-wound analog watch to the battery-run digital watch were in how the time was set, there would be little difficulty. The problem is that new technology has allowed us to add more functions to the watch: the watch can give the day of the week, the month, and the year; it can act as a stop watch (which itself has several functions), a countdown timer, and an alarm clock (or two); it has the ability to show the time for different time zones; it can act as a counter and even as a calculator. But the added functions cause problems: How do you design a watch that has so many functions while trying to limit the size, cost, and complexity of the device? How many buttons does it take to make the watch workable and learnable, yet not too expensive? There are no easy answers. Whenever the number of functions and required operations exceeds the number of controls, the design becomes arbitrary, unnatural, and complicated. The same technology that simplifies life by providing more functions in each device also complicates life by making the device harder to learn, harder to use. This is the paradox of technology.

The paradox of technology should never be used as an excuse for poor design. It is true that as the number of options and capabilities of any device increases, so too must the number and complexity of the controls. But the principles of good design can make complexity manageable.

In one of my courses I gave as homework the assignment to design a multiple-function clock radio:

You have been employed by a manufacturing company to design their new product. The company is considering combining the following into one item:

- *AM-FM radio*
- *Cassette player*
- *CD player*
- *Telephone*
- *Telephone answering machine*
- *Clock*
- *Alarm clock (the alarm can turn on a tone, radio, cassette, or CD)*
- *Desk or bed lamp*

The company is trying to decide whether to include a small (two-inch screen) TV set and a switched electric outlet that can turn on a coffee maker or toaster.

Your job is (A) to recommend what to build, then (B) to design the control panel, and finally (C) to certify that it is actually both what customers want and easy to use.

State what you would do for the three parts of your job: A, B, and C. Explain how you would go about validating and justifying your recommendations.

Draw a rough sketch of a control panel for the items in the indented list, with a brief *justification and analysis of the factors that went into the choice of design.*

There are several things I looked for in the answer. (Figure 17.15 is an unacceptable solution.) First, how well did the answer address the real needs of the user? I expected my students to visit the homes of potential users to see how their current devices were being used and to determine how the combined multipurpose device would be used. Next, I evaluated whether all the controls were usable and understandable, allowing all the desired functions to be oper-

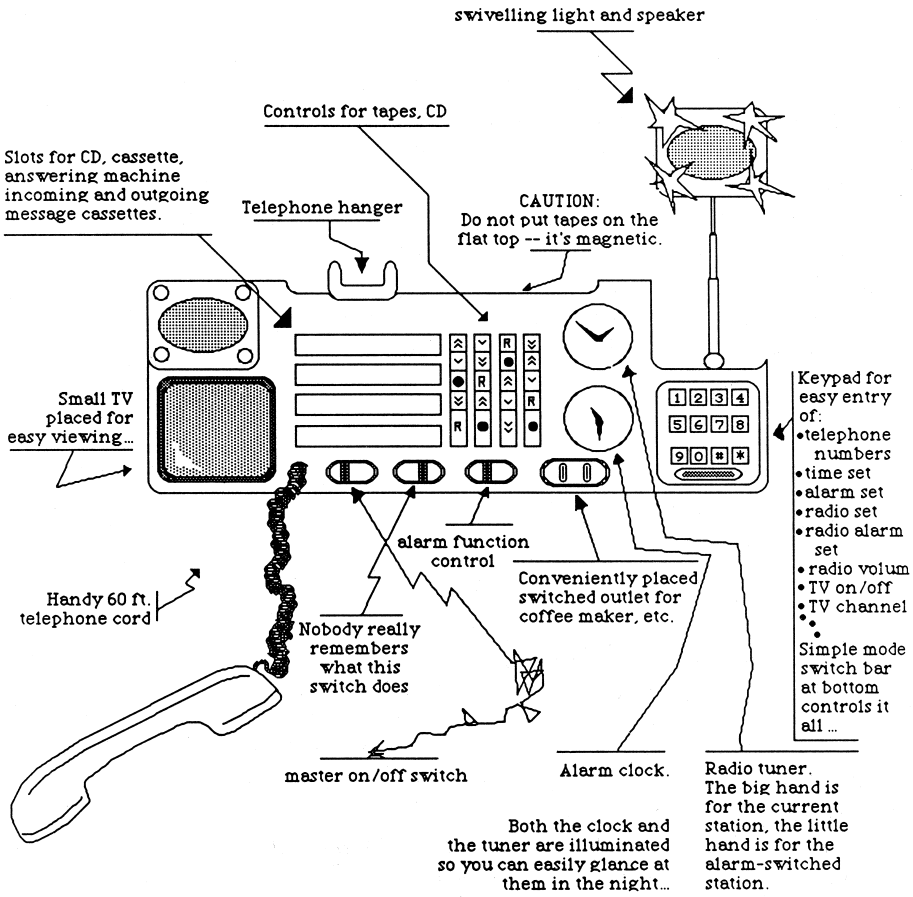


Figure 17.15
Possible Solution to My Homework Assignment. Completely unacceptable. (Thanks to Bill Gaver for devising and drawing this sample.)

ated with minimum confusion or error. Clock radios are often used in the dark, with the user in bed and reaching overhead to grope for the desired control. Therefore the unit had to be usable in the dark by feel only. It was not supposed to be possible to make a serious mistake by accidentally hitting the wrong control. (Alas, many existing clock radios do not tolerate serious errors—for example, the user may reset the time by hitting the wrong button accidentally.) Finally, the design was expected to take into account real issues in cost, manufacturability, and aesthetics. The finished design had to pass muster with users. The point of the exercise was for the student to realize the paradox of technology: added complexity and difficulty cannot be avoided when functions are added, but with clever design, they can be minimized.

Notes

1. Reprinted by permission of the *Wall Street Journal*, © Dow Jones & Co., Inc., 1986. All rights reserved.
2. W. H. Mayall (1979), *Principles in design*, 84.
3. The notion of affordance and the insights it provides originated with J. J. Gibson, a psychologist interested in how people see the world. I believe that affordances result from the mental interpretation of things, based on our past knowledge and experience applied to our perception of the things about us. My view is somewhat in conflict with the views of many Gibsonian psychologists, but this internal debate within modern psychology is of little relevance here. (See Gibson, 1977, 1979.)
4. D. Fisher & R. Bragonier, Jr. (1981), *What's what: A visual glossary of the physical world*. The list of the eleven parts of the sink came from this book. I thank James Grier Miller for telling me about the book and lending me his copy.
5. Biederman (1987) shows how he derives the number 30,000 on pages 127 and 128 of his paper, "Recognition-by-components: A theory of human image understanding," *Psychological Review, 94*, 115–147.
6. I thank Mike King for this example (and others).
7. More complex systems have already been successfully built. One example is the speech message system that recorded phone calls for later retrieval, built by IBM for the 1984 Olympics. Here was a rather complex telephone system, designed to record messages being sent to athletes by friends and colleagues from all over the world. The users spoke a variety of languages, and some were quite unfamiliar with the American telephone system and with high technology in general. But by careful application of psychological principles and continual testing with the user population during the design stage, the system was usable, understandable, and functional. Good design is possible to achieve, but it has to be one of the goals from the beginning. (See the description of the phone system by Gould, Boies, Levy, Richards, & Schoonard, 1987.)

References

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115–147.

Fisher, D., & Bragonier, R., Jr. (1981). *What's what: A visual glossary of the physical world*. Maplewood, NJ: Hammond.

Gibson, J. J. (1977). The theory of affordances. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing*. Hillsdale, NJ: Erlbaum.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

Gould, J. D., Boies, S. J., Levy, S., Richards, J. T., & Schoonard, J. (1987). The 1984 Olympic message system: A test of behavioral principles of system design. *Communication of the ACM, 30*, 758–769.

Mayall, W. H. (1979). *Principles in design*. London: Design Council.

Norman, D. A. (1986). Cognitive engineering. In D. A. Norman & S. W. Draper (Eds.), *User centered system design: New perspectives on human-computer interaction*. Hillsdale, NJ: Erlbaum.

Chapter 18

Distributed Cognition

Donald A. Norman

Modern commercial airplanes fly with two or three people in the cockpit. One, who sits in the front left-hand seat, is the captain, the person in charge. A second pilot, the first officer, sits in the front right-hand seat. In older aircraft, a third person, the second officer or flight engineer, sits sideways just behind the first officer, facing a panel of controls and displays on the wall of the cockpit. The captain and first officer usually alternate jobs, one flying the airplane during one leg of the trip, the other flying during the next leg, so they also designate themselves by the labels "pilot flying" and "pilot not flying."

The two pilots sit in front of a large panel, the captain's side largely duplicating the first officer's side, with a large control wheel—something like the steering wheel of a car—in front of each pilot. The two wheels are linked, so that whenever one pilot turns one wheel or moves it forward or back, the other wheel follows along. In between the pilots is another set of instruments for controlling the engines, radios, and flaps. These instruments and controls are used by both pilots, so there is only one set (see figure 18.1).

Control rooms—whether the cockpit of a commercial airliner or an industrial plant—tend to contain great big controls. In power plants, there are huge electrical switches, huge meters that display the state of the plant. Because there may be thousands of controls and displays—in one nuclear power plant that I studied, there were an estimated four thousand controls and displays—the rooms are huge, as large as a small house. Several people will normally be monitoring the controls, depending upon the plant and the activity taking place. Large controls in spacious control rooms are the norm in industry. I have seen similar displays and controls in large ships, chemical processing plants, manufacturing plants, and even the control room for one of the lines of the Paris Metro.

The first thought that strikes the modern scientist looking at the controls is that they seem quaint and old-fashioned. When I first saw a nuclear power control room, I was also struck by the thought: "Why on earth does it have to be so big?" Sure, once upon a time, you probably needed a big wheel to turn the rudder of a ship or to operate the control surfaces of an airplane. Once you needed big electrical switches to control all the current that passed through them. You needed big meters and indicators so that they could be seen as the operators walked up and down in the control room. But today none of this is necessary. Most modern equipment is controlled remotely. It is no longer

From chapter 6 in *Things That Make Us Smart* (Reading, MA: Addison-Wesley, 1993), 139–153. Reprinted with permission.

Figure 18.1
Cockpit of the Boeing 747-400 Airplane. This is a modern "glass" cockpit, with most of the mechanical gauges of older aircraft replaced by computer-controlled displays. The captain sits in the left chair, the first officer in the right. The control wheels (just in front of each pilot's chair) are yoked—connected so that both move together. Most of the instruments and controls in front of each pilot are duplicated for the other. Many of the instruments and controls in the center are shared. (Photograph courtesy of Boeing Commercial Airplane Group.)

necessary for the control wheel to actually turn the rudder or operate the airplane's wing surfaces. The large lever that controls the landing gear of an airplane no longer actually moves the gear up and down. No, the controls simply send signals to electric or hydraulic motors that do the actual movement.

It would be entirely possible to take the huge room filled with controls for the power plant or the large control panels of the ship and commercial airplane and put them on a small computer: Show the displays on a couple of colorful computer screens and operate the controls with a simple keyboard, a small switch panel, and the ability to turn things on and off just by touching appropriate areas of the screen. Not only could one do this, but it has been done: Excellent examples of these displays and controls exist in the research and development laboratories for all these industries and, for that matter, in the game world, where one can often purchase excellent simulations of the real devices as games for the home computer—simulations that are good enough to be the model for a real control.

The new technologies seemingly eliminate the need for the large controls required by the old-fashioned mechanical technology. The lesson has not been lost on designers. The new airplanes from Airbus have no control wheels. In-

stead, the two pilots each have small joysticks, not unlike the ones used for computer games. The captain has a small joystick on the left side of the airplane, controlled with the left hand. The first officer has a small joystick on the right side of the airplane, controlled with the right hand. Unlike the control wheels of traditional aircraft, which are interconnected so that one turns along with the other, these two joysticks are independent. They could both be used at once, without either pilot noticing. The airplane's computer decides which one to follow.

Taking this idea a step further, the American National Aeronautics and Space Administration (NASA) has a prototvpe advanced cockpit in its simulator facilities at the Ames Research Center in California that has a typewriter keyboard in front of each pilot. Make those controls smaller and you could free up a lot of space for the pilots. Then you could even enlarge the windows, so they could see better out the windows as they were flying.

It turns out, though, that those big outdated rooms, those large outdated controls, offer many benefits. The benefits are important to the distributed nature of the job. Although many modern plants and most airplanes can be controlled by a single person, when problems arise, it is valuable to have several people around, the better to share the work load, the better to make decisions.

The critical thing about doing shared tasks is to keep everyone informed about the complete state of things. The technical term for this is *situation awareness*: Each pilot or member of the control team must be fully aware of the situation, of what has happened, what is planned. And here is where those big controls come in handy.

When the captain reaches across the cockpit over to the first officer's side and lowers the landing-gear lever, the motion is obvious: The first officer can see it even without paying conscious attention. The motion not only controls the landing gear, but just as important, it acts as a natural communication between the two pilots, letting both know the action has been done. In fact, the motion helps the captain remember that the task was done: Flip a bunch of tiny switches and it might be hard to remember whether the landing-gear switch was flipped. Lean over and pull down a huge lever and the memory of that muscle movement is distinct and retained. The same with the control wheels: When one pilot moves the controls, the other pilot knows it. Automatically, naturally, without any need for talking.

Now consider the two small joysticks used in the all-electronic Airbus aircraft. Many who study aviation are very concerned about the unintended side effects of these sticks: The natural communication between the two pilots is lost. There is no way for one pilot to tell whether the other pilot is controlling the airplane except by asking. There have already been instances of confusion, in which each pilot thought the other was controlling the plane, whereas in fact neither was. In other cases, both thought they were in control at the same time. Neither situation is good. The same problems do occur with the control wheels, but those problems can be detected rapidly, for the movement (or lack of movement) of the wheels presents large visible cues. Moreover, it is easy to look over at the other pilot and check for a hand on the wheel or other large controls; it is not so easy to see whether the hand is on the small, side-mounted joystick.

The need for communication and synchronization of actions among members of a team is a very subtle phenomenon. The large mechanical controls and the resulting large control rooms required people to move around a lot as they did their tasks. As a result, a lot of communication was shared, but invisibly, accidentally, without people really being aware that it was happening. Nobody realized just how important this was to the smooth operation of the system until it went away.

A similar situation was observed when my colleague Edwin Hutchins studied the navigation procedures used in large ships in the United States Navy. Members of the navigation team communicated with one another through telephone handsets, so each could hear what the others were saying. The person taking bearings of landmarks from the port side of the ship could hear the person taking bearings off the starboard side. The chief and the plotters heard everything. Periodically there were errors. The bearer takers were instructed to look for inappropriate landmarks, or the readings were reported or recorded wrong. When equipment broke down, manual corrections had to be applied to the readings given by the magnetic compasses, and during the initial stages of the breakdown, when everyone was under some stress and time pressure, more errors were made.

The normal response of the cognitive scientist to the babble of voices over the telephone sets and the prevalence of error is to try to simplify things, to get rid of the error. Maybe the telephone lines should be connected individually to each member of the team so they wouldn't have to listen to all that irrelevant stuff from the other people. The error rate certainly ought to be worked on: Error can't be a good thing. Wrong.

Hutchins showed that the shared communication channel and, especially, the shared hearing of the errors was critical to the robustness and reliability of the task. A navigation team is a permanent fixture of a ship, but the individual members of the team are continually changing. At any one time, the team is composed of individuals who vary in skill from novices to accomplished experts. The shared communication keeps them all informed. The shared listening to the errors and the corrections acts as an informal, but essential, training program, one that is operating continually and naturally, without disrupting the flow of activity. In fact, two different kinds of people are being trained simultaneously. It is obvious that the person who made the error is being trained. It is not so obvious that the rest of the crew is also learning from the event: The less experienced crew members learn by hearing of the error and listening to the correction activities; the more experienced crew members are learning how to train, noting what kinds of error correction and feedback are effective, what kinds are disruptive. Over the years, as the shipmates change which part of the task they are responsible for, as some members leave and new ones join, this shared communication channel, with its shared teaching and correcting process, keeps everyone at a uniformly high level of expertise.

The unplanned properties of the large control rooms that enhance social communication and the training roles played by the detection and correction of errors teach several lessons. The most important deal with the nature of shared work, shared communication. These are subtle activities, and we still know re-

markably little about how this process takes place, about what factors make shared work a pleasant, effective interaction and what factors make it stressful, inefficient, and ineffective.

Many of the essential properties of effective shared action seem to result from "accidental" side effects of the old-fashioned way of doing things. I put the word *accidental* in quotes because I suspect the procedure is not quite as accidental as it might seem, even if it was never consciously designed. That is, over years of experience, the procedures for performing these tasks have gone through a process of natural evolution from their original form to their current shape. Over time, a long sequence of minor changes would occur, each modifying procedures in a small way. Changes that were effective would be apt to stay; changes that were detrimental would be apt to die away. This is a process of natural evolution, and it can lead to remarkably efficient results, even if nobody is in charge, even if nobody is aware of the process.

It is dangerous to make rapid changes in long-existing procedures, no matter how inefficient they may seem at first glance. New technologies can clearly provide improvements over old methods. The old-fashioned control rooms are indeed old-fashioned. Many of their properties, even the ones people grow fond of, are accidental by-products of the technology and may even be detrimental to the task. New technologies can indeed make life more enjoyable and productive. The problem is, it isn't always obvious just which parts are critical to the social, distributed nature of the task, which are irrelevant or detrimental. Until we understand these aspects better, it is best to be cautious.

Natural, smooth, efficient interaction should be the goal of all work situations. Alas, natural interaction is often invisible, unnoticed interaction: We don't know it is there until we remove it, and then it may be too late. We do know that communication is important, however. Listening to the chatter of air traffic controllers turns out to keep pilots informed about all the other airplanes along the route: Replacing this chatter with computer messages sent only to the relevant airplanes destroys this critical aspect of situation awareness, even while giving the benefit of more accurate, less confusing messages. In a similar fashion, replacing the office clerk who delivers mail from department to department with a computer-controlled robot also destroys one channel of communication among departments. Automating factory control or forms processing can also hinder the informal communication processes among workers that allow productive, unofficial decision paths to develop within a company.

The human side of work activities is what keeps many organizations running smoothly, patching over the continual glitches and faults of the system. Alas, those inevitable glitches and faults are usually undocumented, unknown. As a result, the importance of the human informal communication channels is either unknown, unappreciated, or sometimes even derided as an inefficient and obstructive, non–job-related activity.

Eventually, the natural process of evolution will work even upon the latest of technologies. The problem is that in the meantime, if we are too precipitous in making change solely because it is possible, we are apt to run into difficulties. When these difficulties occur in commercial aircraft or large industrial plants, the results can be tragic.

Disembodied Intelligence

The sciences of cognition have tended to examine a disembodied intelligence, a pure intelligence isolated from the world. It is time to question this approach, to provide a critique of pure reason, if you will. Humans operate within the physical world. We use the physical world and one another as sources of information, as reminders, and in general as extensions of our own knowledge and reasoning systems. People operate as a type of distributed intelligence, where much of our intelligent behavior results from the interaction of mental processes with the objects and constraints of the world and where much behavior takes place through a cooperative process with others.

In the research areas studied by experimental psychologists, linguists, and workers in the field of artificial intelligence, thought and understanding are assumed to take place with little or no hesitation, little or no error, and little or no doubt. Scientists make these assumptions in order to simplify their task. "After all," they will state, "the phenomena we are studying are so complex that it is essential to look at them first without all those other complicating factors. Then, after we have understood the isolated case, we can move on to the more realistic and complex situations." The problem with this point of view is that the so-called simplification may be making the task more difficult.

With a disembodied intellect, isolated from the world, intelligent behavior requires a tremendous amount of knowledge, lots of deep planning and decision making, and efficient memory storage and retrieval. When the intellect is tightly coupled to the world, decision making and action can take place within the context established by the physical environment, where the structures can often act as a distributed intelligence, taking some of the memory and computational burden off the human. To give one example: Linguists are continually worried about the amount of ambiguity that exists within language. A huge amount of scientific research has gone into developing schemes for understanding and trying to minimize this ambiguity. But the ambiguity almost always results from the analysis of single, isolated sentences: in real situations, where several interacting people deal with real events, the sentences usually have only one meaningful interpretation. Actually, even when communications are ambiguous, they are usually not perceived as such by either speaker or listener, even though both may have different interpretations of the meaning. It is this lack of perception of ambiguity that is important, and it derives from the communicative, social nature of language, something that is entirely missed when the language is studied as isolated, "simplified" printed sentences or utterances, completely abstracted from the real, social setting.

Information in the world can be thought of as a kind of storehouse of data. This has many advantages. The world remembers things for us, just by being there. When we need a particular piece of information, we simply look around, and there it is. Do I need to repair my car? I don't have to remember the exact shape of the part, because when the time comes for me to do the task, the shape is there in front of me. This eases the burden on initial data collection, eases the requirements on learning and memory, and avoids the need for complex indexing or retrieval schemes. Moreover, it guarantees that the values so obtained will be the most timely available at the moment of need.

Of course, it is important to plan ahead, but postponing decisions until the point of action can simplify the thought processes: Many alternatives that would have had to be thought of ahead of time will turn out not to be relevant. Moreover, the physical structures available in the world can then guide the selection of relevant choices.

Approaches to reasoning and planning that rely heavily upon thought, and therefore internal information, run into fundamental problems:

- *Lack of completeness:* In most real tasks, it simply isn't possible to know everything that is relevant.
- *Lack of precision:* There is no way that we can have precise, accurate information about every single relevant variable.
- *Inability to keep up with change:* What holds at one moment may not apply at another. The real world is dynamic, and even if precise, complete information were available at one point in time, by the time the action is required, things will have changed.
- *A heavy memory load:* To know all that is relevant in a complex situation requires large amounts of information. Even if you could imagine learning everything, imagine the difficulty you'd have finding the relevant stuff just when it is needed. Timely access to the information becomes the bottleneck.
- *A heavy computational load:* Even if all the relevant variables were known with adequate precision, the computational burden required to take them all properly into account would be onerous.

The negative side of this is that these world-based decisions must be made and actions must be taken quickly, which can cause oversimplification and incomplete analysis. We all know that actions taken in haste are often wrong actions. With time pressures, there is limited opportunity to consider alternatives or to reflect upon all of the consequences. Clearly, we need to plan ahead, but not to follow those plans rigidly. We need to respond to the situation, to be flexible in the face of unexpected occurrences, to change our activities as the world dictates.

In the World, Impossible Things Are Impossible

The world has an important property: In the real world, it is not possible to do actions that are not possible. This sounds trivial and obvious, but it has some profound implications when we move into the artificial world of cognitive artifacts. Thus it is certainly not trivial to those who write computer programs that mimic the world. Much of the effort of writing programs that simulate the world must be devoted to ensuring that the simulation cannot do impossible things.

I have flown in extremely sophisticated simulations of aircraft, ones that barely could be distinguished from the real thing. These professional simulators were constructed from real cockpits, they vibrated and sounded like real planes, and moved about two meters in all directions so they could simulate most of the body sensations. And when you looked out the window, you saw the appropriate sights. Yes, the planes behaved just right. But I once flew in a

727 simulator around the streets of San Francisco, a commercial airline pilot at the controls, flying around the Transamerica building. Oops, we flew through the Transamerica building. Not even a tremble. Once we dove into the ground a close to supersonic speeds. Those of us in the cockpit felt somewhat nauseous: Our minds expected sights, sounds, and movements that did not occur. The computer simulator just kept going. Buildings, walls, even the ground are just numerical and graphic abstractions: To a simulator, there is nothing impossible about being 1 meter below the ground.

Suppose a programmer of computer games developed an exploratory game with Harjimé, the protagonist, wandering through the halls of the enchanted castle. Writing the part of the program that controls Harjimé isn't all that difficult, nor is the part of the program that simulates the castle. Want to simulate the room with the hidden treasure? Just draw in the locations of the walls, furniture, secret keys and panels, and the hidden door. But making the simulation work would be a complex task. The hard part is to make sure that Harjimé doesn't walk through the physical objects in the room. If Harjimé picks something up and then puts it down somewhere else, the programmer has to worry about whether there is a supporting structure at the new location, or if the object will fall, tilt, or slide. Harjimé's movements would also have to be carefully monitored to make sure there was always a supporting floor or surface. Harjimé couldn't move up or down unless there was always a suitable support (but he would have to move up, down, or around when he encountered stairs, ramps, furniture, and elevators). Although the task of programming Harjimé and the castle could be given to novices, the task of programming the interaction of the two is complex and difficult enough that it would tax even the experts: How quickly the program could actually execute would be determined by how well it could compute the necessary constraints and interactions.

The point is that in the real world, the natural laws of physics allow only the appropriate things to happen. There is no need to compute whether you are walking through a wall: You simply can't do it. In the artificial world of computer simulation, much of the computational effort goes into the part that results from the artificiality of the situation.

It has long been noted that in dreams, people are free of the constraints of everyday life. We can visualize doing things that are impossible in the real world. Ah, the freedom of dreams, the fantasies released. The impossible actions of dreams might be ways by which people satisfy their fantasies. But they might also result from the impoverished programming power of the human mind.

Suppose, just suppose, that the wonderfully creative fantasies of our dreams are artifacts, accidents of the fact that our minds can't quite handle the computational job of doing accurate simulation. A dream, after all, is a simulation of human action within a simulated environment. The simulation program is executed within the human mind—a disembodied mind, however, for the sense organs are inhibited and the voluntary muscle system inoperative. Consider what it would take to run this simulation properly. The people and objects would have to be created and their actions determined. The environment would have to be created. And finally, the interactions among all the

objects and people and the environment would have to be simulated, which means continual checks to ensure that two objects don't pass through one another, that the force of gravity worked properly, and that impossible actions did not take place. It would be a complex programming job, and one that put enormous computational demands upon the brain.

How much easier to simplify the computations. Let objects pass through one another. Let gravity work in inadequate ways. Free yourself from the constraints of the real world and you reap enormous benefits. The effort is much reduced and the result much more intriguing. Now the human interpretive system can go to work on the products of its own (inadequate?) simulation. It frees up the creative spirits, allows us to contemplate the impossible, amuses and entertains, and creates the industry of dream analysis. Imagine, all these side effects result from simplifying the computational load of the simulation.

Why Accuracy Is Not Always Important

In the days of oral tradition, before reading and writing were widespread, it used to be common for storytellers to go from village to village, telling stories, passing news from one place to another. Here what was important was style and content. These storytellers were famed for prodigious feats of memory, for they could often tell stories that lasted for hours to an enthralled audience. The stories were all memorized. And when modern scholars studied the few remaining storytellers in the few remaining preliterate cultures, they were proudly told how accurate the memory of these storytellers was.

But when the stories were tape-recorded and compared, any particular story varied tremendously from telling to telling. Where was the accuracy? One telling might be twice as long as another. Yet to a villager who had heard both renditions, they were identical, except that perhaps one was better than the other.

To the listener and teller both, word-for-word accuracy was unnecessary. The very notion is not even understood by a completely oral culture. It is only with the advent of writing and tape recorders that we care about such things. It is only the scholar who carefully writes every word of one telling and compares it, word for word, with the next. As for the rest of us, in our normal group settings and activities, who notices, or even cares?

The storytellers didn't memorize the stories, at least not in the sense of the word-for-word learning that we call memorizing today. Basically, the story framework was learned, plus formulas for filling out the phrases and color. The fact that the tales were told in poetry helped, for this put further constraints on the possible wording: The story had to follow the story line, fit the well-known formulas, and fit the meter and rhyme of the poem. The storytellers were able to construct the story anew for each telling, varying it according to the characteristics of the audience. But still, it was the same story, and the listener who heard it once when it lasted one hour would insist it was the same thing as heard the previous week or year when it might have lasted two hours. It was the same story, except for the details of the telling. The fact is, we are social, interacting people, always alert to interpretations, meanings, and reasons. We

need stories and context. Who cares whether the details vary? Who cares whether there is word-for-word accuracy? That is simply not important for everyday life.

Human memory is organized around the important things in life: the excitement, the meaning, and the experience itself. Word-for-word accuracy is simply not important, and it is difficult to accomplish. However, this is no longer true in today's technological world. Great accuracy is required. Lawyers watch every step. Machines are sensitive to every deviance. We are forced to use memory in ways not natural to its evolutionary biological history. And so we must turn to artifacts.

Beware: Using artifacts—technology—to help overcome the frailty of human memory may move us in undesired directions and swamp us with excessive amounts of excessively precise information. The question "What can technology do to help?" is almost always the wrong question. Sure, we can devise technological solutions to the problem. Maybe we can invent small, powerful computers that will remember for us, computers small enough to be available at all times. If not computers, tiny voice recorders small enough to be worn on the wrist. But once we start thinking this way we become trapped in an ever-lengthening chain of technology dependence that in turn forces us to deal with an ever-increasing load of detailed information. Because we can't readily grasp all of this, we will need to devise additional technology to aid us, putting us even more at the mercy of our machines. The whole solution is wrong because the problem is wrong. The correct approach is to structure the world so that we do not have to remember such mindless trivia. Then the question of technological aids would never have been asked. No "solution" would have been necessary.

This is the lesson from the preceding sections of the chapter. Those large control rooms may be unnecessary today, but in changing them, we must be sensitive to the social communication that they afford: Changing the equipment may accidentally destroy the informal communication channels that make work proceed smoothly, synchronized among a group of workers without the need for direct verbal communication.

In airplanes and navy ships, shared communication may at first seem unnecessary, exposing people to irrelevant messages. But, the messages carry information about the activities of others, information that at times is essential to the smooth synchronization of the task or, as in the case of the ship navigators, information that serves as an efficient training device for the entire crew, regardless of their level of expertise. People are effective when they work in a rich, varied environment. A disembodied intelligence is deprived of rich sources of information.

Finally, some aspects of technology expose us to demands for accuracy and precision that are of little importance to normal life. Nonetheless, we have altered our lives to give in to the machine-centered focus on high accuracy, even where accuracy is not critical. Our goal should be to develop human-centered activities, to make the environment and the task fit the person, not the other way around.

PART X

Music Cognition

Chapter 19

Neural Nets, Temporal Composites, and Tonality

Jamshed J. Bharucha

In this chapter, I outline a framework in which aspects of cognition can be understood as the result of the neural association of patterns. This approach to understanding music cognition originates with Pitts and McCulloch (1947) and Deutsch (1969). Subsequent advances (e.g., Grossberg, 1970, 1972, 1976; Rumelhart and McClelland, 1986) enable us to understand how these neural associations can be learned. Models based on these mechanisms are called neural net models (also connectionist models or parallel distributed models).

Neural net models have a number of properties that recommend them as models of music cognition. First, they can account for how we learn musical patterns through exposure and how this acculturation influences our subsequent perception of music. Second, their assumptions are either known or plausible principles of neuroscience. Third, they shed light on the observation (Terhardt, 1974) that aspects of pitch and harmony involve the mental completion (or Gestalt perception) of patterns. Fourth, they are capable of recognizing varying shades of similarity and are therefore well suited to modeling similarity-based accounts (e.g., Krumhansl, 1990) of tonality or modality. Finally, they can discover regularities in musical styles that may elude formal music-theoretic analysis (Gjerdingen, 1990).

Section I of this chapter deals with neural representation, and Section II deals with neural association and learning.

I. Neural Representation

A. Frequency Tuning of Neurons

Many neurons, particularly sensory neurons, are highly selective in their response. For example, there are neurons in the auditory system that respond selectively to specific bands of frequencies. Within this band, there is usually a frequency to which the neuron responds maximally (called the *characteristic frequency*). For the purpose of the present analysis, the signal to which a neuron responds maximally may be called a *feature*, and the neuron itself may be called a *feature detector*. A neuron that has a characteristic frequency may be called a *frequency detector*. A feature detector responds progressively less strongly to signals that are increasingly dissimilar. This relationship is given by its *tuning curve*. The left-hand panel of figure 19.1 shows a schematic tuning curve of a frequency detector.

From chapter 11 in *The Psychology of Music*, 2d ed., ed. D. Deutsch (San Diego, CA: Academic Press, 1999), 413–440. Reprinted with permission.

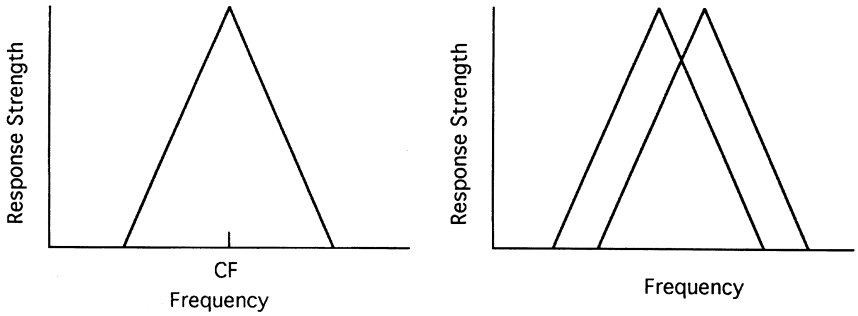


Figure 19.1
Left: Tuning curve. A frequency detector responds most strongly to a particular frequency—its characteristic frequency (CF)—and less so to frequencies farther away. Right: Coarse coding achieved by overlapping tuning curves of frequency detectors with different characteristic frequencies.

Frequency detectors can be found at almost all major stages in the auditory system, including the inner ear (Tasaki, 1954), the auditory nerve (Russell & Sellick, 1977), the cochlear nucleus (Rose, Galambos, & Hughes, 1959), the inferior colliculus (Semple & Aitkin, 1979), the medial geniculate body (Gulick, Gescheider, & Frisina, 1989) and the auditory cortex (Merzenich, Knight, & Roth, 1975). In all those structures, neurons seem to be arranged *tonotopically*, that is, systematically in order of characteristic frequency. Although most of the studies reporting tonotopy have involved animals, recent positron emission tomography studies have revealed tonotopic frequency tuning in humans at the cortical level (Lauter, Hersovitch, Formby, & Raichle, 1985). Many of the representations used in this chapter are tonotopic, although only their tuning, and not tonotopy per se, is computationally relevant. The networks that operate on these representations would function equivalently if the neurons were arranged randomly while preserving their tuning.

It may at first seem odd to think of frequencies as features, because frequency is a continuous dimension that is infinitely dense, that is, between the lowest and highest frequencies, we can detect an infinite number of frequencies. Yet the brain represents this continuum with a finite set of detectors with characteristic frequencies at discrete points. We are unaware of the gaps between the characteristic frequencies because each frequency detector responds to a broad band of frequencies around its characteristic frequency, in accord with its tuning curve, and the response band overlaps with those of other neurons (right-hand panel of figure 19.1). Any given frequency thus activates an entire family of inner hair cells to various degrees, with the strongest response coming from the neuron whose characteristic frequency is closest to the sounded frequency. This form of representation is called *coarse coding*. Coarse coding enables a perceptual dimension to be denser than the array of neurons used to perceive it. (Cones in the retina have only three different characteristic wavelengths, yet we can discriminate hundreds of different colors).

Coarse coding permits the listener to assimilate small tuning differences to broad musical categories (such as semitones) while also permitting us to detect fine degrees of mistuning. The former is enabled because the broadening of the

peaks creates substantial overlap between the representations of patterns that differ only slightly in their tuning. The latter is enabled because the maxima of the peaks are unchanged and can be recovered if necessary by sharpening the peaks through lateral inhibition.

B. Abstract Feature Tuning

Although frequency detectors have been the most widely studied feature detectors in the auditory system, evidence exists for detectors of more abstract features. Pantev, Hoke, Lütkenhöner, and Lehnertz (1989) argue that the tonotopic representation in the primary auditory cortex is of pitch, not frequency. Weinberger and McKenna (1988) have found feature detectors for contour. Frequency detectors must therefore map onto higher order neurons in such a way as to extract pitch and contour from complex spectra. This suggests a hierarchy of feature detectors: elementary features are detected at the sensory periphery, and entire patterns of these features are detected by abstract feature detectors, which in turn form patterns that are detected by even more abstract feature detectors. This conception of neural architecture has already received strong support for the visual system (Hubel & Wiesel, 1979; Linsker, 1986; Marr, 1982).

Deutsch (1969) suggested how feature abstraction might occur in a neural net. For example, if frequency detectors whose characteristic frequencies are an octave apart connect to the same neuron, and if no other frequency detectors connect to this neuron, then it is effectively an octave-equivalent frequency detector. The circuits Deutsch proposed anticipate the circuits that develop automatically as a result of learning, although these methods were not available at that time.

The neural connections that make a unit an abstract feature detector may in many cases have developed through evolution, in which case they are innate. Yet it seems obvious that humans are capable of learning new patterns, and if abstract feature detectors are necessary for pattern learning (as most neural net models tacitly assume), then humans must be capable of acquiring abstract feature detectors through learning. We understand how this can be done (Fukushima, 1975; Grossberg, 1970, 1972, 1976; von der Malsberg, 1973), and in the case of music, it seems reasonable to adopt a presumption of learning.

Neural net models assume an array of units whose feature-detecting properties are given. The network then acquires either new associations or new abstract feature detectors through learning. These two types of learning are commonly referred to as pattern association and self-organization, although the latter can be thought of as a special case of the former. Both types of learning are surveyed in Section II.

The most commonly assumed feature detectors in models of music that learn are pitch or pitch-class (i.e., octave-equivalent pitch) detectors whose tuning is spaced at semitone intervals (e.g., Bharucha, 1987a, 1987b, 1988, 1991; Bharucha & Todd, 1989; Laden & Keefe, 1989; Leman, 1991; Sano & Jenkins, 1989, Todd, 1988). We already have evidence of pitch detectors in the brain (Pantev et al., 1989). Pitch-class units can be postulated on the assumption of a circuit like the one proposed by Deutsch (1969), which if not innate, can be learned by the self-organization of harmonic spectra. The semitone spacing of their

tuning is an interesting issue that is beyond the scope of this chapter. For present purposes, it suffices to think of the set of pitch or pitch-class detectors with semitone spacing as a subset of the more dense array that we know to exist. Semitone spacing is thus not an additional postulate in these models, because if the dense array were used, the feature detectors between the semitones would simply not play much of a role (Bharucha, 1991).

When modeling the learning of musical sequences that are invariant across transposition, an *invariant pitch-class representation* is appropriate (Bharucha, 1988, 1991). A complete invariant pitch-class representation would have 12 units corresponding to the 12 pitch-class intervals above the tonic, which may be referred to as Units 0 through 11 for tonic through leading tone, respectively. Note that in an invariant pitch-class representation, a melody is conceived not as a series of melodic intervals but as a series of scale degrees (i.e., intervals between each note and the tonic). The mapping from pitch class to invariant pitch class can be accomplished by a circuit as described in Section II,D.

Gjerdingen (1989b) uses an invariant pitch-class representation that is restricted to the major diatonic scale (*do, re, me*, etc.), with two extra units representing sharp and flat, respectively. Although the scale degrees can be mapped from pitch-class representations (Section II,D), it is not clear how units representing sharp and flat are acquired.

C. Activation

Precisely how a neuron responds to the features to which it is tuned varies. Neurons in the auditory nerve with characteristic frequencies below 4000 Hz tend to spike at preferred intervals of time that correspond to integer multiples of one cycle of the characteristic frequency. The probability distributions of these interspike intervals are extremely provocative, given the simple integer ratios they generate quite naturally, and suggest a timing code for pitch (Cariani & Delgutte, 1992), harmony (Tramo, Cariani, & Delgutte, 1992), and possibly rhythm.

Beyond the auditory nerve, little evidence exists for timing as a coding strategy. In the cochlear nucleus (the first junction from the auditory nerve to the brain) and beyond, a neuron typically fires more rapidly the more intense the tone, or the closer the tone is to its characteristic frequency. The more rapidly a neuron fires, the more pronounced is its effect on neurons to which it is connected, by virtue of the temporal summation that occurs at the receiving neuron. Firing rate is thus taken to be the measure of response strength for most neurons, and frequencies are represented by a *spatial code*—which neurons are firing and how strongly—rather than by a timing code.

Most neural net models of pitch and tonality use spatial codes, and time enters the coding scheme by changing a spatial code over time. In a spatial code, each neuron has some response strength or *activation* at any given time. Activation is an abstract term for response strength and entails no commitment to an underlying mechanism, although firing rate and temporal summation lend neurophysiological plausibility to the postulation of activation as a theoretical construct for modeling cognitive phenomena. The term activation is also used in models in which the units are postulated as cognitive rather than neu-

ral units, as in spreading activation network models (J. R. Anderson, 1983; Collins & Quillian, 1969). The underlying mechanism could well be the response strength of a group of neurons rather than an individual neuron (Hebb, 1949). For this reason we shall use the term *unit* instead of *neuron* in the context of a model, reserving the latter term for units that are known to be individual neurons.

Numerous mechanisms have been used to model changes in activation over time. These include phasic versus tonic responses, oscillating circuits, temporal composites, and cascaded activation. The first two will be summarized briefly in this section, and the last two will be covered more extensively in later sections.

In the cochlear nucleus and beyond, tones elicit both *phasic* and *tonic* responses. A phasic response is a response to change (usually the onset or offset of a tone); a tonic response is sustained throughout the duration of a tone. Most neurons in the cochlear nucleus show a strong phasic response to the onset of a tone, followed by a weaker tonic response over the sustained portion of a tone (Kiang, 1975). Some neurons (the so-called octopus cells) show only a phasic response to onsets. This enhancement of onsets may serve to draw attention to a new event and may play a role in segmenting the musical stream. Phasic activation can account for the salience of harmonic rhythm (Bharucha, 1987a). Some neurons switch from tonic to phasic as intensity increases (Gulick, Gescheider, and Frisina, 1989); thus both onset and intensity are to some extent coded by a phasic response. This helps explain how chord changes can compete with high-intensity percussive sounds in establishing the meter. (In rock music, for example, the highest intensity percussive sound is often on a weak beat, and it is presumably the chord changes that establish the meter).

Activation can be modulated cyclically over time by oscillatory circuits. These circuits give the activation an isochronous pulse, and have been used by Gjerdingen (1989a) to model meter. Phasic activation and oscillatory circuits are consistent with the idea, proposed by Jones (1989), that meter is temporally focused attention: phasic responses draw attention to the onsets of tonal changes, and oscillatory circuits are implementations of attention via the modulation of activation.

D. Vector Spaces as Formal Depictions of Neural Representations

The representations and computations in a neural net can be understood in terms of linear algebra. Consider, for simplicity, an environment with exactly 2 features, f_1 and f_2. Each possible pattern in the environment consists of some combination of intensities of these two features and can therefore be represented as a point in a Cartesian space whose axes are f_1 and f_2. Clearly, patterns that contain only one of the features will be represented as points along one of the axes. For simplicity, we shall limit the discussion in this chapter to the first quadrant, that is, to feature intensities that are either zero or positive, never negative.

Pattern p, depicted on the left in figure 19.2, contains both features, but f_1 is about twice as intense as f_2. Pattern q lies on the straight line passing through the origin and p. All points on this line (in the first quadrant) represent patterns that contain the two features in the same proportion of intensities. These

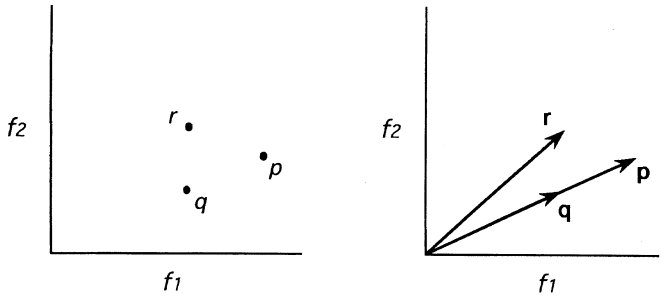


Figure 19.2
Left: Patterns in Cartesian space. Right: The same patterns in vector space.

patterns are essentially the same, but with different intensities, because what defines a pattern is the *relative* intensities of its features. Pattern *r*, in contrast, is a distinct pattern because it contains the features in different proportions.

A vector space is an improvement over a Cartesian space for the depiction of patterns because it makes explicit the equivalence of patterns that vary only in intensity. If instead of points, we draw arrows from the origin to each point, we get vectors **p**, **q**, and **r** (depicted on the right in figure 19.2). Vectors that are oriented in the same direction but have different lengths (e.g., **p** and **q**) are collinear and represent the same pattern with different intensities. Vectors that are oriented in different directions (e.g., **p** and **r**) represent different patterns. The more divergent the directions in which two vectors point, the more dissimilar are the patterns. Without the benefit of a visual depiction of a vector space (as when the vector has more than three dimensions), one can tell if two vectors are collinear by seeing if the multiplication of one vector by a scalar yields the other (Appendix A).

In the perception of pitch and tonality, we are typically concerned with differences between patterns rather than with differences in absolute intensity of the same pattern. Neural nets are responsive to the differences between patterns, and the absolute intensities play a minimal role. This is just one of several reasons why neural net models hold promise for understanding pattern perception, and why vector spaces are promising conceptualizations of these models. In some models, the absolute intensities are ignored altogether in order to simplify the computation. In Grossberg's models, for example, all vectors are normalized so that the sum of the squares of the intensities of all the features equals one. This ensures that all vectors have unit length (i.e., all the vector arrowheads terminate at a unit circle or unit hypersphere centered at the origin).

E. Composite Patterns

The *addition* of vectors yields a *resultant vector* (Appendix B) and is equivalent to superimposing patterns. The resultant represents a pattern that is more similar to each of the original patterns than they are to each other. Graphically, the resultant vector is the diagonal of the parallelogram formed by the two summed vectors, as shown in the left-hand panel of figure 19.3. It makes a smaller

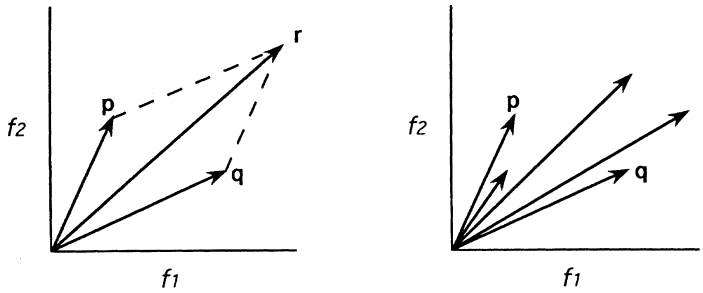


Figure 19.3
Left: Vector addition. r is the sum of p and q. Right: Composite patterns resulting from linear combination of vectors. All vectors within the angular sweep of p and q are composites of them.

angle with each of the original vectors than they make with each other. In more than two dimensions, the resultant vector is the diagonal of the hyperparallelogram formed by the summed vectors.

The resultant vector can be thought of as a *composite*. Composites are superimposed patterns. They have some of the properties of prototypes, being perceived as more familiar than any of the original patterns themselves (Metcalfe, 1991; Posner & Keele, 1968). It is fruitful to expand the notion of composite to include all vectors that result from linear combinations of the original vectors. A linear combination is the addition of vectors that may have scalar coefficients. In a two-dimensional vector space, all the vectors that lie between two given vectors are linear combinations—or composites—of them (right-hand panel of figure 19.3). We can then think of a composite as the result of adding intensity-scaled versions of several vectors.

F. Compositing Patterns over Time

As a piece of music unfolds, patterns can be composited over time by the accumulation of activation, creating a *temporal composite memory*. Suppose, for example, that the features of interest are pitch classes. When a musical sequence begins, the pattern of pitch classes that are sounded at time t_0 constitutes a vector, $\mathbf{p}_0$, in 12-dimensional pitch-class space. If at a later time, t_1, another pattern of pitch classes is sounded, represented by vector $\mathbf{p}_1$, a composite, $\mathbf{c}_1$, covering a period of time ending at t_1, can be formed as follows:

$$\mathbf{c}_1 = s_1\mathbf{p}_0 + \mathbf{p}_1,$$

where s_1 $(0 \leq s_1 \leq 1)$ is the persistence of $\mathbf{p}_0$ at t_1. When yet another set of pitch classes is heard at time t_2, the resulting composite, $\mathbf{c}_2$, is:

$$\mathbf{c}_2 = s_2\mathbf{c}_1 + \mathbf{p}_2.$$

When the nth set of pitch classes is sounded at t_n, the composite, $\mathbf{c}_n$, is

$$\mathbf{c}_n = s_n\mathbf{c}_{n-1} + \mathbf{p}_n,$$

where s_n, the persistence of $\mathbf{c}_{n-1}$ at t_n, brings about the diminution of the represented salience of $\mathbf{c}_{n-1}$ as it recedes into the past. It controls the relative weighting of the most recent pattern, $\mathbf{p}_n$, relative to the patterns that came

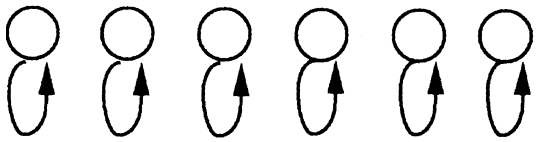


Figure 19.4
Implementing a temporal composite with links from each unit to itself.

before. When $s_n = 0$, there is no temporal integration—no memory except for the most recent event. When $s_n = 1$, events at different points in time are compressed into a composite on equal terms.

Evidence for the persistence of tonal activation is clear. A chord sounded for as short a duration as 50 msec can prime a subsequent chord even if they are separated by as much as 2.5 sec of silence (Tekman & Bharucha, 1992). *Priming* refers to the automatic (i.e., robust and difficult to suppress) expectation for a target event following a context and is measured by the extent to which the context increases the speed and accuracy with which the target is perceptually processed.

The implementation of s_n in a neural net varies among models that have either explicitly or implicitly adopted a temporal compositing representation for music. In the MUSACT model (Bharucha, 1987a, 1987b), the persistence of a previously heard pattern decays exponentially over time. If d $(0 \leq d \leq 1)$ is the decay rate (i.e., the proportion by which activation decreases per unit time), and if t is the number of time intervals since the last event, then:

$$s_n = (1 - d)^t.$$

Although the duration of each time interval controls the temporal resolution of the representation, d determines the length of the temporal window over which information is being integrated.

Although activation in the MUSACT model is strictly phasic, it would be reasonable, in future modeling efforts, to hypothesize a strong phasic response to the onset of a sound followed by a weaker tonic response. This amounts to the continuous formation of new composites, even during the duration of a tone, and is most easily modeled by computing new composites at small and equal time intervals. A sequence would thus be represented as a composite of a series of vectors at the ends of successive time intervals. Some time intervals would include event onsets and others would not. If the time intervals are sufficiently small, this scheme could capture some of the nuances in pitch that are lost when music is represented as a score of notes. Temporal composites can also be explored as a way to represent the spectral flux dimension of timbre.

Temporal composites with small time intervals derive some plausibility from temporal summation in the nervous system. Zwislocki (1960, 1965) found that thresholds for detecting tones show a trade-off between duration and intensity (as would be predicted by a temporal composite) and suggests that a combination of decay and summation of neural activity can account for this.

Some models implement persistence by linking each unit to itself, as shown in figure 19.4 (Bharucha, 1988; Bharucha & Todd, 1989; Todd, 1988, 1989). Each

unit thus activates itself in proportion to its own current activation and the strength of the link. The strength of the link will be referred to as its weight. If the weight is w_i, then:

$$s_n = w^t.$$

This scheme is functionally identical to decay, but is easier to implement.

An alternative postulate to decay is *interference*. Interference is the displacement (or reduction in retrievability) of items in memory by more recently perceived items. Interference seems to occur in both short-term memory (Waugh & Norman, 1965) and long-term memory (Bjork, 1989), but in the present context we are concerned primarily with short-term memory. Interference in short-term memory is typically attributed to a capacity limitation in attention or activation (J. R. Anderson, 1983; Shiffrin, 1975), although specific interactive effects have been noted. In the context of network models, interference is the reduction in activation of some units due to an increase in activation of others. Interference assumes that the total amount of activation among a given set of units is limited, so that activation caused by the currently perceived event comes at the expense of activation caused by earlier events. Interference is usually implemented by introducing inhibition. Gjerdingen (1990) has used what amounts to an interference mechanism, in which the activation of perceived events persists until inhibited by more recent events.

Whether decay or interference accounts for forgetting in short-term memory is a debate that goes back to the very beginnings of cognitive psychology (see Neisser, 1967), and there is evidence for both (Reitman, 1974; Waugh & Norman, 1965). The notion of persistence in a temporal composite, as outlined earlier, is agnostic as to the mechanism and its implementation.

A temporal composite has also been adopted unwittingly by Parncutt and Huron (1993)—although not in the form of a neural net—to account for key tracking data. Parncutt and Huron refer to their representation as *echoic memory*. Echoic memory is an auditory sensory memory that persists for several seconds, after which it is lost unless attended to (Darwin, Turvey, & Crowder, 1972; Neisser, 1967). Echoic memory enables us to relate what we are hearing at this very moment to what we have just heard. It permits us to maintain a temporal window wide enough to recognize a dynamic sound or parse a phrase.

The persistence s_n may also be influenced by segmentation cues—factors that cue the listener to chord changes or to boundaries between groups, motifs, phrases, or other segments. Segmentation cues could include phasic signals for chord changes (see earlier) or any number of pitch, timing, and timbral cues in either the composition or the performance (Bregman, 1990; Lerdahl & Jackendoff, 1983; Palmer, 1989). A segmentation cue would cause s_n to be small, so that a fresh temporal composite can be started for the next segment.

G. Tonal and Modal Composites

A temporal composite of a pitch-class representation may be called a *tonal* composite, and a temporal composite of an invariant pitch class representation may be called a *modal* composite. Tonal composites that integrate information between chord changes represent the chords that have been either played or implied, and can account for aspects of the implication of harmony by melody.

The corresponding modal composites represent chord functions. Tonal composites over longer durations represent keys, and modal composites represent modes. If metrical bias is added, say in the form of pulsing activation (Gjerdingen, 1989a), then a tonal or modal composite would encode an interaction between tonal/modal and metrical information.

If persistence is large and activation is phasic, a tonal composite roughly represents the probability distribution of pitch classes in a segment of music. Krumhansl (1990) has shown that distributions of pitch classes are strongly correlated with empirically determined key profiles of Krumhansl and Kessler (1982). A tonal composite with large persistence is thus a representation of the hierarchy of prominence or stability of pitch classes as determined by their frequency of occurrence in a segment of music. With both tonic and phasic activation, the tonal composite would represent something between the distribution of occurrences of pitch classes and the distribution of durations of pitch classes, both of which are highly correlated with key profiles (Krumhansl, 1990). What Parncutt and Huron (1993) have attributed to echoic memory is a tonal composite of pitch class; their demonstration that some of Krumhansl's probe-tone results can be modeled by such a memory is support for the existence of tonal composites as representations.

Although the distribution of pitch classes in a piece of music has a substantial influence on our perception of the relative stability of pitch classes, long-term representations of structural regularities (sometimes referred to as *schemas*) also exert an influence. In a cross-cultural study by Castellano, Bharucha, and Krumhansl (1984), Western and Indian subjects heard a rendition of a North Indian *rāg* and then judged how well a probe tone fit with the preceding segment. Probe-tone ratings were obtained for all 12 pitch classes following each of 10 *rāgs*. For both Western and Indian listeners, the probe-tone ratings were highly correlated with the distribution of total durations of pitch classes in the segment, consistent with a temporal composite representation. However, the Indian subjects showed an influence of prior exposure to the underlying scale or *thāt*, whereas the Western subjects did not. In a multiple regression analysis, a significant contribution to the regression was made by the distribution of durations for both groups of subjects, but only for the Indian subjects was a significant contribution made by the membership of pitch classes in the underlying *thāt*. This latter variable—*thāt* membership—was assessed by using a 12-element vector of binary elements representing the presence or absence of each pitch class. The contribution of this vector to the multiple regression for the Indian subjects suggests that whereas the responses of the Western subjects were based entirely on the distribution of pitch classes in the most recently heard segment, the responses of the Indian subjects were based also on prior knowledge of which pitch classes are typically present (i.e., the hierarchy of stability of tones was internalized).

This prior knowledge is implicit, schematic, and acts like a cultural filter. Implicit knowledge can be studied by priming. In a priming task, a target stimulus is presented following a context (prime stimulus), and subjects are instructed to make a designated true/false decision about the target. If the speed and accuracy with which the decision is made are greater following context C_1 than following context C_2, then C_1 primes the target more than C_2

primes the target. Priming thus reveals the extent to which one stimulus evokes another. Priming tasks are well suited to studying music cognition because of their robustness across levels of formal expertise. Because of the premium on speed, there isn't time for musical experts to use analytical strategies; and if the true/false decision is one that novices can make, priming can reveal associations that the novice may be unable to express verbally.

Priming studies demonstrate that musical events that typically co-occur in a musical culture become mentally associated. For Western listeners, for example, chords that have high transition probabilities in Western music prime each other, even though they may share no frequencies (Bharucha, 1987b). For Indian listeners, tones that typically co-occur in a particular *thāt* prime other tones in the *thāt* (Bharucha, 1987b).

Tonal and modal composites can account for these results if they are encoded for later retrieval. Pitch classes or invariant pitch classes that co-occur in a temporal composite can become mentally associated if the composite is stored in memory. The long-term encoding of composited information can be accomplished by neural nets that adjust the connections between units, and is the subject of Sections II,A and II,B.

Although it may seem that the temporal integration in tonal or modal composites results in a complete loss of information about the serial order of events, serial order can indeed be recovered, as needed for the recognition and performance of pieces of music, if the context is unambiguous. Section II,C deals with the long-term encoding and recovery of sequences using modal composites.

II. Neural Association and Learning

This section deals with how a neural net can learn temporal composite patterns so that they function as schemas and as sequential memories. (Some of these mechanisms may be limited to modal composites for most of the population but may extend to tonal composites for absolute pitch possessors.)

A neural net (equivalently, a connectionist or parallel distributed network) consists of units connected by links. Links have weights associated with them, representing the strengths of the connections between units. The *net input* to a unit at any given time is a weighted sum of activations received through the links that connect to it (*spatial summation*), integrated over time (*temporal summation*). We will deal only with spatial summation first and then introduce temporal summation later. Spatial summation can be modeled as follows. The net input, net_j, to unit j, is

$$net_j = \sum_i w_{ij} a_i$$

where w_{ij} is the weight associated with the link from unit i to unit j, and a_i is the activation of unit i. The summation is over all units i that connect to j.

Links may be unidirectional or bidirectional, or may conduct different kinds of information in different directions. Although synapses in the brain are typically unidirectional, departures from strict unidirectionality in a neural net model are not neurophysiologically implausible, because separate sets of synapses could underlie different directions of information flow.

Sections II,A–C deal with mechanisms that enable a network to learn patterns by finding an appropriate set of interconnections. In each case, initial constraints on connectivity are specified, and the learning mechanism determines how these existing connections are strengthened or weakened. The initial constraints usually take the form of layers of units, with connections from one layer to the next; this architecture is supported by the layered organization of the cerebral cortex. The mechanisms for changing the connection strengths derive from Hebb's (1949) hypothesis that when two connected neurons are active simultaneously or in close temporal succession, the connection between them is strengthened so that eventually the activation of one will lead indirectly to the activation of the other. The models discussed in Section II,A—self-organizing models—use this so-called Hebbian learning in close to its original form. The models discussed in Section II,C learn by error correction and use a modified version of Hebbian learning: the connection between two units changes as a function of the activation of one unit and the error registered by the other. The models discussed in Section II,B—autoassociators—can use either Hebbian learning or error correction, although the latter enables them to learn many more patterns and to distinguish grades of similarity.

A. Encoding Temporal Composites: Abstract Feature Detectors or Category Units

Sensory neurons are stimulated directly by energy external to the organism. Their tuning characteristics are a consequence of their inherent transducing properties and are innately fixed. For example, the inner hair cells convert mechanical deformation of the basilar membrane into neural signals. In contrast, neurons beyond the sensory periphery are stimulated by other neurons that connect to them, not by the environment directly. Their tuning characteristics are based on the pattern of stimulation they receive from other neurons. These can be called abstract feature detectors or category units (because they encode entire categories).

The connectivity that achieves this can be learned by a class of learning models called self-organizing neural nets (Grossberg, 1970, 1972, 1976; Rumelhart & Zipser, 1985; von der Malsburg, 1973). Grossberg's models, the earliest and most fully developed of this kind, have been used to model the acquisition of auditory categories in music (Gjerdingen, 1989b, l990) and speech (Mitra, 1993). Although a detailed description of this model would require a chapter in itself, it is possible to capture the essence of self-organizing models rather simply. (Most of the specifics of Grossberg's theory deal with ensuring the stability of the learned categories, and the stability of human categories is an open question.)

The top panel of figure 19.5 shows a layer of units (input units), with preexisting tuning characteristics, connected to another layer (category units). The category units are in a winner-take-all configuration, which is common in the brain: the most active unit in such a configuration has the effect of decreasing the activation of the other units and boosting its own activation. A pattern presented to the network will activate the input units with the corresponding features (filled circles). The ensuing activation of the category units depends on the weights on the links. One of the category units will win (filled circle in the

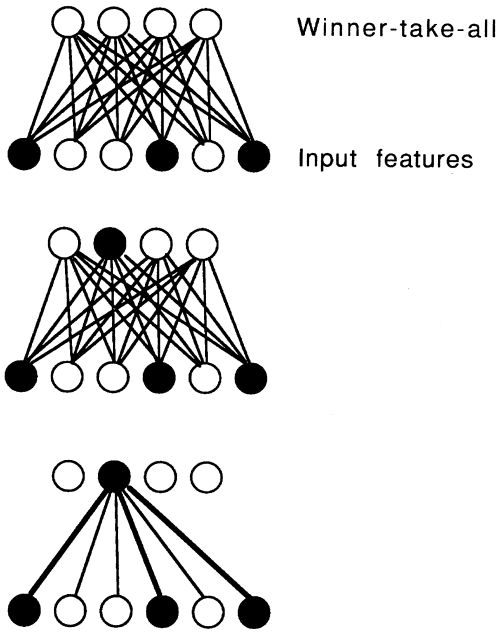


Figure 19.5
Self-organization. The winning category unit gets to learn. Links to it from highly active input units are strengthened.

middle panel), and the weights on the links feeding into the winning unit change by Hebbian learning. The links from strongly active input units are strengthened (bottom panel). Self-organizing mechanisms have the further requirement that the links to the winner from weakly active input units are weakened. The winner is on its way to becoming a feature detector or category unit for the entire input pattern. Similar patterns will activate this unit more strongly, and dissimilar patterns will activate this unit more weakly, than before learning.

Self-organization can be visualized in terms of vector spaces. Consider a network with two input units, f_1 and f_2, and indefinitely many units in a second layer. The units in the second layer are available to become abstract feature detectors and may be called category units. Each category unit has two links feeding into it, one from each input unit. The weights on these links can be plotted as vectors (solid lines in figure 19.6) in two-dimensional feature space; these are *weight vectors*—each category unit has a weight vector. A pattern presented to the network can be plotted as a vector (dashed line) in the same space. The weight vector that is closest in angle to the pattern vector represents the category unit that has responded most strongly to the pattern and is therefore the most likely candidate for an abstract feature detector for that pattern. The weights of this unit are changed so as to move the weight vector closer to the pattern vector. The closer the weight vector moves to the pattern vector, the more strongly this unit will respond to that pattern, that is, the more it develops the tuning characteristics for that pattern.

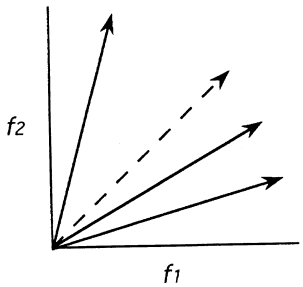


Figure 19.6
The weight vector (solid line) with the smallest angle to the activation vector (dashed line) represents the winning category unit. Learning consists of making the angle even smaller.

If the input is a tonal composite, this procedure will lead to the formation of abstract feature detectors for typical composite patterns. The chord and key units in the MUSACT model (Bharucha, 1987a, 1987b) are thus a direct and mandatory consequence of self-organization. Pitch classes that co-occur within a tonal composite that spans the duration of a chord (explicit or implied) will become associated via the chord detectors that form. Chords that co-occur within a tonal composite that spans a piece or a key segment will become associated via the key detectors that form. After these associations are sufficiently strong, hearing one chord will lead to expectations for other chords that co-occur in the same composite, because activation flows from one chord unit to parent key units and down to the other chord units in the same composite. The MUSACT model suggests how the graded activation of chords, mediated by the multiplicity of their parent keys, and the priming data that provide evidence of this, can be explained by this process (Bharucha, 1987a, 1987b).

Adopting an input representation that is essentially a temporal composite of invariant pitch-class units, Gjerdingen (1989b) exposed a self-organizing network to works of early Mozart. The network developed categories units (abstract feature detectors) for sequential patterns that characterize the style of the corpus.

B. Encoding Temporal Composites through Autoassociation

Consider a network in which each unit is connected by unidirectional links to every other unit and to itself. This network can learn to encode patterns through autoassociation, that is, by associating them with themselves (J. A. Anderson, 1970, 1972; J. A. Anderson, Silverstein, Ritz, & Jones, 1977). Why would one wish to associate patterns with themselves? Neural net autoassociators have a remarkable property: If after learning a set of patterns, an incomplete or degraded version of one of the learned patterns is presented to the network, it will be completed or filled in by the network. Pattern completion is a general principle of perception that enables us to recognize objects that are partially masked or occluded and to perceive them as unbroken wholes, with the concomitant risk of error or illusion. Terhardt (1974) has argued that many auditory phenomena are examples of this aspect of Gestalt perception.

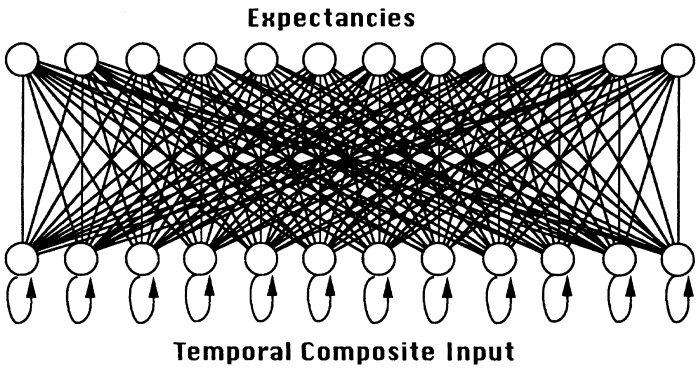


Figure 19.7
An autoassociator with pitch-class units or invariant pitch-class units as input and expectation. The input units constitute a temporal composite, and each input unit is connected to each expectation unit.

Aspects of tonality can be thought of as pattern completion. The residual effect of prior exposure found in the responses of Indian listeners hearing Indian *rāgs* (Castellano, Bharucha, and Krumhansl, 1984) is evidence of this: a temporal composite of the underlying mode accounted for variance over and beyond the variance accounted for by the probability distribution of pitch classes in the segment. The segment seems to have activated an internal representation of the mode, which in turn elaborated or filled out the percept. More direct evidence comes from the priming of an important tone missing from a *rāg*, based on the remaining tones (Bharucha, 1987b). Subjects' responses in these studies seem to reflect a composite of the distribution of pitch classes actually heard during the experiment and an internal representation of the pitch classes that typically occur in that context. This should not be surprising at all, because the literature in perception is filled with examples of top-down processing, that is, the influence of context-dependent expectations based on prior experience.

In an autoassociator, the links between units serve to excite units whose pitch classes co-occur and inhibit units whose pitch classes do not. This requires just the right combination of weights on these links, because two pitch classes may co-occur in one key or mode and not in another. Although this may seem like an impossible standard for this tangled network to meet, a simple learning mechanism can lead to this result.

For the purposes of illustration, it is useful to duplicate the units and think of one copy as representing the stimulus that is actually heard and the other as representing expectations that are triggered by this stimulus. Figure 19.7 shows an array of pitch class units that represents, as a temporal composite, what has been heard (the input) and another array of pitch-class units that represents expectations based on what has been heard. Each input unit feeds into each expectation unit. A learning mechanism called the *delta rule* enables the weights to adjust themselves so each of a number of patterns presented repeatedly to the input units will reproduce itself at the expectation units. The delta rule derives from the *perceptron* developed by Rosenblatt (1962).

According to the delta rule, as adapted shortly, the weights are assumed to be random initially, representing a naive network. A tonal composite of a key as input will initially result in a random pattern of expectations. This random pattern of expectations is compared with the input pattern, and the weights are changed so as to reduce the disparity. The weight change is incremental; each time the network generates expectations in response to a tonal composite, the weights change slightly so that the next time that tonal composite (or one similar to it) is encountered, the expectations will more closely approximate the input.

If a_i is the activation of input unit i, and a_e is the activation of expectation unit e, then the disparity is called the *error signal* (δ_e), and is simply the difference:

$$\delta_e = a_i - a_e.$$

If w_{ie} is the weight on the link from i to e, then the change (Δw_{ie}) in the weight is simply the activation of i times the error signal at e, scaled by a constant, ε ($0 \leq \varepsilon \leq 1$) that represents the learning rate:

$$\Delta w_{ie} = \varepsilon a_i \delta_e.$$

The learning rate determines the extent to which a single experience can have a lasting effect.

The patterns are presented repeatedly to the network until the error signal is smaller than some criterion amount for all expectation units in response to all patterns. Rosenblatt (1962) proved that, with this learning rule, a network with two sets of units, one feeding into the other, will eventually be able to find a solution (to any given degree of precision) if one exists. For an autoassociator using the delta rule, a solution exists for any set of input vectors that are *linearly independent* of each other. A vector is linearly independent of a set of vectors if it cannot be obtained by any combination of scalar multiplication and addition of the other vectors, that is, it is not a *composite* of any of the others. Tonal composites for the 12 major keys are linearly independent of each other, and modal composites for the Church modes are linearly independent of each other. This is indeed a powerful system, because it can learn all these patterns in the same set of links.

Two modes that have the same invariant pitch classes, albeit with different probability distributions (e.g., major and natural minor), are not linearly independent; this network would learn them as one pattern that is a composite of the two. This is not a limitation of this scheme for modeling music cognition, however, because it has never been suggested that tonal or modal composites capture all features of music. These models can be expanded to include any number of features that may discriminate two modes by the way in which they are used. We have restricted ourselves to pitch classes or invariant pitch classes in the present chapter only because we must begin somewhere and it behooves us to understand something well before bringing all possible factors into play.

After learning, the model can be tested for its vaunted ability to complete degraded patterns or to assimilate similar patterns to learned ones. An autoassociator was fed tonal composites, in each of the 12 major keys, representing the average probability distributions of works by Schubert, Mendelssohn, Schumann, Mozart, Hasse, and Strauss (as reported by Krumhansl, 1990, p. 68).

After learning, the network was presented with temporal composites that were similar to but not identical to one of the learned composites. For example, the network was presented with a composite in which the pitch classes D, E, F, G, and A were equally active (i.e., the vector 0,0,1,0,1,1,0,1,0,1,0,1). The network recognized that this pattern was more similar to the C major composite than to any other and significantly activated all and only the diatonic pitch classes among the expectation units, including C, which was missing in the input.

With invariant pitch-class units, an autoassociator can learn modes. Bharucha and Olney (1989) presented an autoassociator with binary modal composites of 10 North Indian *rāgs*. After the network learned them, it was tested with incomplete patterns. *Rāg Bhairav*, for example, contains the invariant pitch classes: 1,1,0,0,1,1,0,1,1,0,0,1 (which in C major would be C,D♭,E,F,G,A♭,B). When the network was presented with all the tones except the second scale degree (D♭), all the scale degrees were activated among the expectation units, including the missing second scale degree. The network generated these expectations with a much smaller set of tones: the third, fourth, sixth, and seventh scale degrees were sufficient to suggest *Bhairav*.

C. Learning Sequences

The expectations that derive from the above system are *schematic*—expectations for classes of events rather than specific event tokens—based on familiarity with a musical culture (Bharucha & Todd, 1989). They are also not sequential, but rather represent global states or backgrounds against which the actual sequences of events are heard. Yet tonal or modal composites can also serve as the basis for encoding specific sequences. A memory for specific sequences, when activated by appropriate context, generates *veridical expectancies*—the cues that enable us to anticipate or recognize the next event in a familiar piece and that underlie our ability to perform from memory.

The system shown in figure 19.8 is a sequential memory that serves this function and has the added bonus that while it learns specific pieces it also learns something about the sequential regularities—sequential schematic expectancies—of the style. The architecture is similar to that of the autoassociator in figure 19.7 in that there is a set of input units and expectation units. The input feature space is given more dimensions to include additional features that play a role in cueing one's memory for the continuation of a sequence. Candidates for these additional features are contour, timbre, aspects of rhythm and, because human memory is highly contextual, even aspects of the extramusical context that might cue memory; these additional context units could conceivably receive input from systems far afield from the auditory system.

The system works by generating an expectation for the next event in a sequence, based on a temporal composite of the sequence thus far. As each new event is heard, it adds to the composite, and the new composite generates an expectation for the following event. The units in the middle, unlabeled, layer of figure 19.8 are called *hidden units*. They are necessary if the system is to be able to learn the full range of possible transitions in musical sequences. Each hidden unit computes a nonlinear, monotonically increasing function, as do neurons: the more strongly activated a neuron, the stronger its response, but because of physical limitations, the response strength asymptotes. One of the more

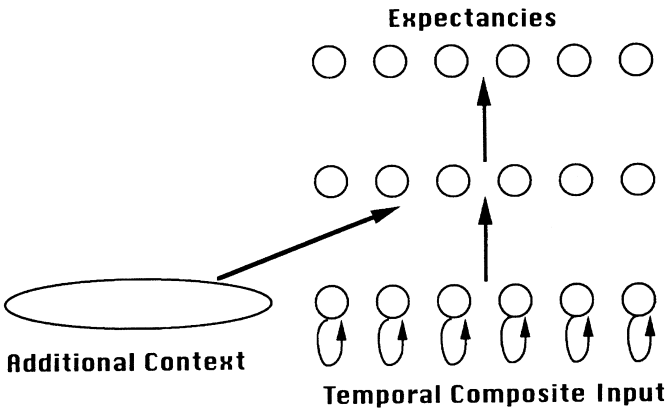


Figure 19.8
A network that learns individual sequences (veridical expectancies) and acquires schematic properties. The "Additional Context" units represent any extratonal information that may be encoded as context. The arrows between groups of units represent a link from each unit of one group to each unit of the other.

commonly used functions in modeling hidden units is the logistic function (figure 19.9).

This nonlinearity of hidden units enables a network to implement mappings from input to expectation that would otherwise be impossible. (There is no advantage to hidden units if they are linear). If the tips of the tonal composite vectors that generate an expectation for pitch class x cannot be clearly separated by a hyperplane from the tips of the tonal composite vectors that generate an expectation for pitch class y, then the expectations are not *linearly separable*. In other words, if there are cases in which similar tonal composites generate different expectations and dissimilar tonal composites generate the same expectation, then the expectations may not be linearly separable. Similar tonal composites tend to generate similar schematic expectancies but not necessarily similar veridical expectancies. This is because composers occasionally use unusual or unschematic transitions that violate (schematic) expectations for aesthetic effect (Meyer, 1956). Problems that are not linearly separable cannot be solved by neural nets without nonlinear hidden units (Minsky & Papert, 1969) or extra assumptions.

We use the logistic function at the expectation units as well because it has the effect of making the activations at the expectation units equivalent to probabilities. The weights in the network are initially random. As a sequence is played, a temporal composite at the input produces a pattern of expectations that is initially random. The network learns by comparing the expectation for the next event with the actual next event when it occurs. Each event thus trains the expectations that attempted to predict it.

The delta rule is adapted for this model as follows. The error signal is scaled by the slope of the logistic function at the expectation unit's current activation level (the derivative of the activation of e with respect to its net input):

$$\delta_e = (t_e - a_e)\frac{da_e}{dnet_e},$$

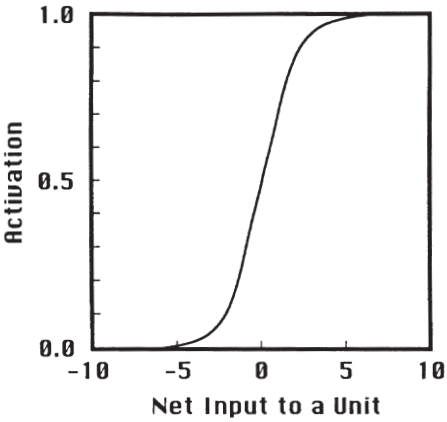


Figure 19.9
Logistic function relating the net input and activation of a unit.

where $t_e = 1$ if event e occurs, 0 otherwise. This has the effect of changing the weight more radically when the unit into which it feeds is uncommitted (in the middle of its activation range). Applying the delta rule to change the weights from the hidden units to the expectation units:

$$\Delta w_{he} = \varepsilon a_h \delta_e,$$

where a_h is the activation of hidden unit h and w_{he} is the weight on the link from hidden unit h to expectation unit e.

The delta rule offers no guidance on how to change the weights from the input units to the hidden units, because the error signal on the hidden units is undefined. The solution, commonly known as *backpropagation* (Rumelhart, Hinton, & Williams, 1986), has dramatically broadened the scope of neural net models in recent years. In the context of the present model, each hidden unit inherits the error of each expectation unit connected to it, weighted by the link between them, and sums these weighted errors. This is again scaled by the slope of the logistic function at the hidden unit's current activation level. Thus the error signal on hidden unit e is:

$$\delta_h = \left(\sum_e w_{he} \delta_e \right) \frac{da_h}{dnet_h}.$$

The delta rule is then applied to change the weights on the links from the input units to the hidden units:

$$\Delta w_{ih} = \varepsilon a_i \delta_h.$$

This model was used to learn sequences of chord functions, using a temporal composite of invariant pitch chord function for input and expectation (Bharucha & Todd, 1989). Figure 19.8 shows six units, representing, in a major key, the tonic, supertonic, mediant, subdominant, dominant, and submediant. Fifty sequences, of seven successive chords each, were generated at random using a priori transition probabilities estimated from Piston's (1978, p. 21) table of

chord transitions. This corpus roughly represents the transition probabilities of chord functions in the common practice era, but contains a small proportion of highly unusual transitions because of the random generation procedure. In order to encapsulate the potentially large number of possible "additional context" features, one additional context unit was assigned to each sequence as a place holder for all the contextual information that might help individuate this sequence.

After repeated presentation of the sequences, the network learned to predict the first event in each sequence in response to the activation of its additional context unit and learned to predict each successive event in response to the temporal composite of chords played thus far plus the activation of its additional context unit. In a performance model, this would enable the performer to play the first event. In a perceptual model, it would enable the listener to recognize whether or not the correct event was played.

After learning, the network was presented repeatedly with two new sequences: one consisted entirely of schematically expected (high-probability) transitions and the other of schematically unexpected (low-probability) transitions. The schematic sequence was learned in fewer presentations. The network adapted more quickly to the sequence that was typical of the corpus than to the sequence that was unusual, even though both were novel. This suggests that the network learned not only the sequences themselves but also the generic or schematic relationships of the style.

The same network therefore contains information about the two types of expectations—veridical and schematic—that usually converge but sometimes diverge. When they diverge, the performer is able to produce, and the listener to recognize, the correct next event while nevertheless experiencing its unexpectedness. The divergence of expectations when an unusual transition occurs in a familiar piece addresses what Dowling and Harwood (1986) refer to as Wittgenstein's puzzle. It also accounts for how expectancy violation, which Meyer (1956) considers central to our aesthetic response to music, can continue to occur in a familiar, overlearned piece.

The network reveals these divergent expectations when the activation of expectation units following the onset of an event is observed over time. We have thus far considered only spatial summation of activation. The buildup of activation in a neuron as an event gets under way is the result of temporal summation. If we consider both spatial and temporal summation, the net input to a unit can be modeled using *cascaded* activation (McClelland, 1979):

$$net_{j,t} = k\left(\sum_i w_{ij}a_{i,t}\right) + (1-k)net_{j,t-\Delta t},$$

where k $(0 \leq k \leq 1)$ restricts the incremental net input in any given time slice, Δt, and the second term of the equation carries over net input form the previous time slice, thereby causing the net input to build up over time.

With cascaded activation, high-probability (schematic) expectations were generated in less time than low-probability expectations (Bharucha & Todd, 1989). Unique expectations, resulting from chord transition that occurred only once in the corpus, took the longest. This is presumably because the net-

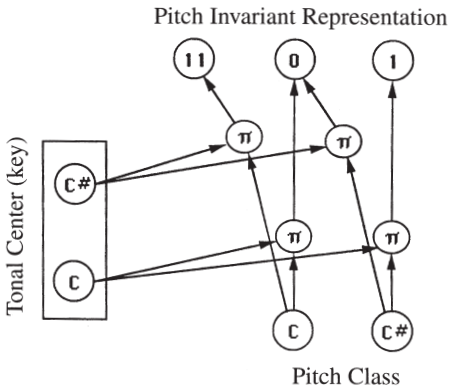


Figure 19.10
A network for transforming a pitch-class representation into an invariant pitch-class representation. Units labeled π gate activation received from pitch class units and tonal center (key) units. (From Bharucha, 1988.)

work develops redundant pathways for the high-probability transitions (e.g., moving to the tonic), leading to the rapid activation of some units. Whether or not the veridical expectancies are also schematic is therefore revealed by the time course of activation. When the unusual transition involves not just a low-probability transition but a move to a low-probability event (such as a secondary dominant), the expectancy violation does not require a cascading explanation; it is trivially accounted for by the disparity between the expectation of the sequential net and the autoassociator.

D. Transpositional Invariance

Pitch-class representations can be transformed into invariant pitch-class representations by a simple gating mechanism (Bharucha, 1988, 1991; McNellis, 1993). This mechanism, shown in figure 19.10, is similar to one developed by Hinton (1981) for object recognition in vision. The units labeled "π" multiply the activation that feeds into them, thereby serving as "AND" gates. The units arrayed vertically on the left are key units from MUSACT, and the units arrayed horizontally at the bottom are pitch-class units. The most active key unit gates the activation from pitch-class units into the pitch invariant representation at the top. If the key is C♯, then C is gated to 11 and C♯ to 0. If the key is C, then C is gated to 0 and C♯ to 1. This can account for the invariance of pitch sequences under transposition.

III. Discussion

Neural net models are not intended to be statements of fact about how the brain is wired. Like all models, they are systematic hypotheses based on available data, and they represent attempts to account for known phenomena and guide further research. Some neural net models may be sufficiently closely tied to known physiology that they serve as hypotheses of actual neural circuitry. For most examples of complex human behavior, however, the precise circuitry

is largely a mystery. We have a large and rapidly growing body of knowledge about the physiology of single neurons and about the functions, and connections between, some macroscopic regions of the brain. We also have a preliminary understanding of how different types of neurons are interconnected within the parts of the brain that are thought to play major roles in cognition, namely, the cerebral cortex and the cerebellum. However, we know little about the specific circuits that underlie specific phenomena, and even when the circuits are known, it is sometimes not clear why the circuit behaves the way it does. Neural net models are attempts to bridge the gap between what we do and do not know.

In the opinion of some authors (e.g., Fodor & Pylyshyn, 1988), the connectionist conception of cognition and the representations that underlie it contrast sharply with rule-based systems, the latter being the hallmark of computer programming languages and the grammars of modern linguistic theory (Chomsky, 1980). Fodor (1975; Fodor & Pylyshyn, 1988) argues, among other things, that the mind is a formal symbol-manipulating device, in which a fairly clear distinction is made between syntactic form and semantic content. Although many of his arguments are specific to the study of language, one incisive argument derives from his critique of connectionism (Fodor & Pylyshyn) together with his theory of the relationship between mental and physical states (Fodor, 1975). Roughly, he contends that connectionist models are merely models of implementation. Because there may be radically different implementations of the same symbolic process (e.g., there may be radically different hardware designs that can implement the same computer program), an understanding of one implementation does not entail an understanding of the formal symbolic process it implements, any more than an understanding of the electrical activity in the circuits of a computer chip entails an understanding of the program it is running.

This is a powerful argument, but a careful analysis is beyond the scope of this chapter. If the argument is correct, connectionist modelers will have to settle for trying to understand how the brain—a *mere* implementation, but what an implementation it is!—implements the formal symbolic processes that we call music cognition. I suspect, however, that although highly trained musicians may use formal symbolic processes together with a host of other processes, the passive processing of music by most listeners is minimally symbolic. What then does one make of rule-based theories of music, such as that of Lerdahl and Jackendoff (1983)? These can be construed as formalizations of constraints on neural processing of music. In other words, either neural nets are implementations of grammars, or grammars are formal descriptions of neural nets. Future research will need to bridge the gap either way.

Appendices

A. Collinearity of Vectors

If $\mathbf{v}$ is a vector and s is a positive scalar, then their product, $s\mathbf{v}$, is a vector that is collinear, that is, will point in the same direction. The multiplication of a vector by a scalar is the vector resulting from multiplying each component by the

scalar. For example, if

$$\mathbf{v} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 3 \\ 1 \end{bmatrix}$$

and $s = 2$, then

$$s\mathbf{v} = \begin{bmatrix} 4 \\ 0 \\ 2 \\ 6 \\ 2 \end{bmatrix}.$$

Division of a vector by a scalar is analogous.

B. Addition of Vectors

The addition of two vectors is the vector resulting from adding their corresponding components. For example:

$$\begin{bmatrix} 3 \\ 0 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 4 \\ 1 \end{bmatrix}.$$

Only vectors with the same number of components can be added.

Acknowledgments

This manuscript was completed while I was a Fellow at the Center for Advanced Study in the Behavioral Sciences in 1993–1994. I am grateful for support from the National Science Foundation (DBS-9222358 and SES-9022192) and for valuable comments from Diana Raffman, Carol Krumhansl, Eugene Narmour, Caroline Palmer, Einar Mencl, Subhobrata Mitra, Mark McNellis, and Denise Vargas.

References

Anderson, J. A. (1970). Two models for memory organization using interacting traces. *Mathematical Biosciences*, *8*, 137–160.

Anderson, J. A. (1972). A simple neural network generating an interactive memory. *Mathematical Biosciences*, *14*, 197–220.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, *84*, 413–451.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Bharucha, J. J. (1987a). MUSACT: A connectionist model of musical harmony. *Proceedings of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Bharucha, J. J. (1987b). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, *5*, 1–30.

Bharucha, J. J. (1988). Neural net modeling of music. *Proceedings of the First AAAI Workshop on Artificial Intelligence and Music*. Minneapolis: American Association for Artificial Intelligence.

Bharucha, J. J. (1991). Pitch, harmony, and neural nets: A psychological perspective. In P. Todd & G. Loy (Eds.), *Connectionism and music*. Cambridge, MA: MIT Press.

Bharucha, J. J., & Olney, K. L. (1989). Tonal cognition, artificial intelligence and neural nets. *Contemporary Music Review, 4*, 341–356.

Bharucha, J. J., & Todd, P. (1989). Modeling the perception of tonal structure with neural nets. *Computer Music Journal, 13*(4), 44–53. (Reprinted in P. Todd & G. Loy, Eds., *Connectionism and music,* 1991, Cambridge, MA: MIT Press)

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 309–330). Hillsdale, NJ: Erlbaum.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Cariani, P., & Delgutte, B. (1992, October). *The pitch of complex sounds is simply coded in interspike interval distributions of auditory nerve fibers*. Paper presented at the annual meeting of the Society for Neuroscience, Anaheim, CA.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General, 113*, 394–412.

Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240–247.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). The auditory analog of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, *3*, 255–267.

Deutsch, D. (1969). Music recognition. *Psychological Review*, *76*, 300–307.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. San Diego, CA: Academic Press.

Fodor, J. A. (1975). *The language of thought*. New York: Crowell.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 2–71.

Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics, 20*, 121–136.

Gjerdingen, R. O. (1989a). Meter as a mode of attending: A network simulation of attentional rhythmicity in music. *Intégral*, *3*, 67–92.

Gjerdingen, R. O. (1989b). Using connectionist models to explore complex musical patterns. *Computer Music Journal, 13*(3), 67–75. (Reprinted in P. Todd & G. Loy, Eds., *Connectionism and music,* 1991, Cambridge, MA: MIT Press)

Gjerdingen, R. O. (1990). Categorization of musical patterns by self-organizing neuronlike networks. *Music Perception, 8*, 67–91.

Grossberg, S. (1970). Some networks that can learn, remember, and reproduce any number of complicated space-time patterns. *Studies in Applied Mathematics*, *49*, 135–166.

Grossberg, S. (1972). Neural expectation: Cerebellar and retinal analogs of cells fired by learnable or unlearned pattern classes. *Kybernetic, 10*, 49–57.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, *23*, 121–134.

Gulick, W. L., Gescheider, G. A., & Frisina, R. D. (1989). *Hearing*. Oxford: Oxford University Press.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hinton, G. F. (1981). A parallel computation that assigns canonical object-based frames of reference. *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 683–685.

Hubel, D., & Wiesel, T. N. (1979). Brain mechanisms of vision. *Scientific American*, *241*, 150–162.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*, 459–491.

Jordan, M. I. (1986). Attractor dynamics and parallelism in a connectionist sequential machine. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Kiang, N. Y. (1975). Stimulus representation in the discharge patterns of auditory neurons. In D. B. Tower (Ed.), *The nervous system* (pp. 81–96). New York: Raven Press.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychology Review*, *89*, 334–368.

Laden, B., & Keefe, D. H. (1989). The representation of pitch in a neural net model of chord classification. *Computer Music Journal, 13*(4), 12–26. (Reprinted in P. Todd & G. Loy, Eds., *Connectionism and music,* 1991, Cambridge, MA: MIT Press)

Lauter, J. L., Hersovitch, P., Formby, C., & Raichle, M. R. (1985). Tonotopic organization in the human auditory cortex revealed by positron emission tomography. *Hearing Research*, *20*, 199–205.

Leman, M. (1991). The ontogenesis of tonal semantics: Results of a computer study. In P. Todd & G. Loy (Eds.), *Connectionism and music*. Cambridge, MA: MIT Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Linsker, R. (1986). From basic network principles to neural architecture. *Proceedings of the National Academy of Sciences, USA, 83*, 7508–7512, 8390–8394, 8779–8783.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

McClelland, J. L. (1979). On the time-relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.

McNellis, M. (1993). *Learning and recognition of relative auditory spectral patterns*. Unpublished honors thesis, Dartmouth College, Hanover, NH.

Merzenich, M. M., Knight, P. L., & Roth, G. L. (1975). Representation of the cochlea within primary auditory cortex in cat. *Journal of Neurophysiology*, *28*, 231–249.

Metcalfe, J. (1991). Composite memories. In W. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 399–423). Hillsdale, NJ: Erlbaum.

Meyer, L. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.

Mitra, S. (1993). *A neural self-organization approach to problems in phonetic categorization*. Unpublished honors thesis, Dartmouth College, Hanover, NH.

Mozer, M. C. (1990). Connectionist music composition based on melodic, stylistic and psychophysical constraints. *University of Colorado Technical Report CU-CS-495–90*.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton, Century, Crofts.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 331–346.

Pantev, C., Hoke, M., Lütkenhöner, B., & Lehnertz, K. (1989). Tonotopic organization of the auditory cortex: Pitch versus frequency representation. *Science*, *246*, 486–488.

Parncutt, R., & Huron, D. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology, 12*, 154–171.

Piston, W. (1978). *Harmony*. New York: Norton.

Pitts, W., & McCulloch, W. S. (1947). How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics*, *9*, 127–147.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353–363.

Reitman, J. (1974). Without surreptitious rehearsal, information in short-term memory decays. *Journal of Verbal Learning and Verbal Behavior*, *13*, 365–377.

Rose, J. E., Galambos, R., & Hughes, J. R. (1959). Microelectrode studies of the cochlear nuclei of the cat. *Bulletin of the Johns Hopkins Hospital*, *104*, 211–251.

Rosenblatt, F. (1962). *Principles of neurodynamics*. New York: Spartan.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 1. Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Vols. 1 & 2. Cambridge, MA: MIT Press.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science*, *9*, 75–112.

Russell, I. J., & Sellick, P. M. (1977). The tuning properties of cochlear hair cells. In E. F. Evans & J. P. Wilson (Eds.), *Psychophysics and physiology of hearing* (pp. 71–84). New York: Academic Press.

Sano, H., & Jenkins, B. K. (1989). A neural network model for pitch perception. *Computer Music Journal, 13*(3), 41–48. (Reprinted in P. Todd & G. Loy, Eds., *Connectionism and music*, 1991, Cambridge, MA: MIT Press)

Semple, M. N., & Aitkin, L. M. (1979). Representation of sound frequency and laterality by units in the central nucleus of the cat's inferior colliculus. *Journal of Neurophysiology*, *42*, 1626–1639.

Shiffrin, R. M. (1975). Short-term store: The basis for a memory system. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory*. Vol. 1. Hillsdale, NJ: Erlbaum.

Tasaki, I. (1954). Nerve impulses in individual auditory nerve fibers of guinea pig. *Journal of Neurophysiology*, *17*, 97–122.

Tekman, H., & Bharucha, J. J. (1992). Time course of chord priming. *Perception & Psychophysics*, *51*, 33–39.

Terhardt, E. (I974). Pitch, consonance and harmony. *Journal of the Acoustical Society of America*, *55*, 1061–1069.

Todd, P. (1988). A sequential network design for musical application. In D. Touretzky, G. Hinton, & T. Sejnowski (Eds.), *Proceedings of the 1988 Connectionist Models Summer School*. Menlo Park, CA: Morgan Kaufmann.

Todd, P. (1989). A connectionist approach to algorithmic composition. *Computer Music Journal, 13*(4), 27–43. (Reprinted in P. Todd & G. Loy, Eds., *Connectionism and music,* 1991, Cambridge, MA: MIT Press)

Tramo, M. J., Cariani, P. A., & Delgutte, B. (1992). Representation of tonal consonance and dissonance in the temporal firing patterns of auditory nerve fibers: Responses to musical intervals composed of pure tones vs. harmonic complex tones. *Society for Neuroscience Abstracts*, *18*, 382.

von der Malsberg, C. (1973). Self-organizing of orientation sensitive cells in the striate cortex. *Kybernetic*, *14*, 85–100.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, *72*, 89–104.

Weinberger, N. M., & McKenna, T. M. (1988). Sensitivity of single neurons in auditory cortex to contour: Toward a neurophysiology of music perception. *Music Perception*, *5*, 355–390.

Zwislocki, J. J. (1960). Theory of temporal auditory summation. *Journal of the Acoustical Society of America*, *32*, 1046–1060.

Zwislocki, J. J. (1965). Analysis of some auditory characteristics. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. New York: Wiley.

Chapter 20

The Development of Music Perception and Cognition

W. Jay Dowling

I. Introduction

An adult listening attentively to a piece of music and understanding it performs an enormous amount of information processing very rapidly. Most of this processing is carried out automatically below the level of conscious analysis, because there is no time for reflective thought on each detail as the piece steadily progresses. This process is closely parallel to what happens when a native speaker of a language listens to and understands a sentence. The elements of the sentence are processed very rapidly—so rapidly that the listener cannot attend individually to each detail, but simply hears and understands the overall meaning. The rapidity of automatic speech processing depends on extensive perceptual learning with the language in question. Similarly, the music listener's facility in grasping a piece of music depends on perceptual learning gained through experience with the music of a particular culture. Further, we can see in the development of language from its earliest stages the predisposition of the child to speak, and the ways in which basic elements of language, already present in infancy, are molded through perceptual learning and acculturation into adult structures (Brown, 1973). Similarly, we can find elements of adult cognitive structures for music in young infants, and can watch them develop in complexity under the influence of culture and individual experience. In both speech and music, then, there are specific patterns of behavior that emerge in infancy that bear the unmistakable stamp of "speech" or "music" behavior. We can trace the elaboration of those incipient speech and music patterns in the course of development.

A point to be emphasized is the ease and rapidity with which adults perform complex cognitive tasks in domains of speech and music familiar to them, and the degree to which that facility depends on prior experience. For example, when the processing of a melody is complicated by the temporal interleaving of distractor notes among the notes of the melody, listeners are more accurate in judging pitches that match familiar, culturally determined norms than those that do not (Dowling, 1992, 1993a). Furthermore, the ability to discern a target melody in the midst of temporally interleaved distractors grows gradually through childhood, and the importance of the culturally defined tonal scheme to the performance of that task grows as well (Andrews & Dowling, 1991). Perceptual learning with the music of a culture provides the listener with a

From chapter 15 in *The Psychology of Music*, 2d ed., ed. D. Deutsch (San Diego: Academic Press, 1999), 603–625. Reprinted with permission.

fund of implicit knowledge of the structural patterns of that music, and this implicit knowledge serves to facilitate the cognitive processing of music conforming to those patterns.

Calling the knowledge amassed through perceptual learning "implicit" indicates that it is not always available to conscious thought. Neither the knowledge base itself nor the cognitive processes through which it is applied are entirely accessible to consciousness (Dowling, 1993a, 1993b). Listeners typically engage in far more elaborate processing than they are aware of. For example, there is evidence that listeners with a moderate amount of musical training encode the diatonic scale-step ("*do, re, mi*") values of the notes of melodies they hear (Dowling, 1986). Yet those listeners are not aware that they are even capable of categorizing melodic pitches according to their scale-step values, much less that they do it routinely when hearing a new melody. Implicit knowledge of Western musical scale structure has accrued over years of experience, and that knowledge is applied automatically and unconsciously whenever the adult listens to music.

This sensorimotor learning undoubtedly has consequences for brain development, as illustrated by Elbert, Pantev, Wienbruch, Rockstroh, and Taub's (1995) demonstration of the enhanced allocation of cortical representation to fingers of the left hand in string players, especially for those who begin study of the instrument before the age of 12. Recent results by Pantev, Oostenveld, Engelien, Ross, Roberts, and Hoke (1998) concerning cortical allocation in processing musical tones tend to confirm this supposition.

In looking at the development of music perception and cognition, one of our goals is to distinguish between cognitive components that are already present at the earliest ages and components that develop in response to experience. We can look at the content of the adult's implicit knowledge base in contrast to the child's. We can also look at the developmental sequence by which the individual goes from the infant's rudimentary grasp of musical structure to the experienced adult's sophisticated knowledge and repertoire of cognitive strategies for applying it.

II. Development

A. Infancy

Over the past 20 years, much has been learned about the infant's auditory world. Researchers have isolated several kinds of changes that infants can notice in melodies and rhythmic patterns, and those results give us a picture consistent with the notion that infant auditory perception uses components that will remain important into adulthood. In broad outline it is clear that infants are much like adults in their sensitivity to the pitch and rhythmic grouping of sounds. This is seen in infants' tendency to treat melodies with the same melodic contour (pattern of ups and downs in pitch) as the same and to respond to the similarity of rhythmic patterns even across changes of tempo. Similarly, we find that in children's spontaneous singing, rhythmic grouping and melodic contour are important determinants of structure and that when children begin singing, their singing is readily distinguishable from speech in terms of its pat-

terns of pitch and rhythm. In both perception and production, we find that the child's cognition of musical patterns contains the seeds of the adult's cognition.

1. Prenatal Experience Even before birth, the infant appears to be sensitive to music, or at least to patterns of auditory stimulation. Research has shown that prenatal auditory stimulation has effects on the infant's behavior after birth. Shetler (1989) has reviewed studies showing that the fetus is responsive to sounds at least as early as the second trimester. Very young infants recognize their mother's voice (DeCasper & Fifer, 1980; Mehler, Bertoncini, Barrière, & Jassik-Gerschenfeld, 1978), and this may derive from neonatal experience with the mother's characteristic patterns of pitch and stress accents. Such an interpretation is plausible in light of the demonstration by DeCasper and Spence (1986) that patterns of a speech passage read repeatedly by their mothers during the third trimester of pregnancy were later preferred by babies. DeCasper and Spence had newborns suck on a blind nipple in order to hear one or another children's story. Children who had been read a story in the womb sucked more to hear that story, while babies who had not been read stories in the womb had no preference between the two stories. Spence and DeCasper (1987) also demonstrated that babies who had been read stories in the womb liked speech that was low-pass filtered (resembling speech heard before birth) as much as normal unfiltered speech, whereas babies who had not been read to did not.

2. Perceptual Grouping Infants' grouping of sounds in the pitch and time domain appears to follow much the same overall rules of thumb as it does for adults. Just as adults segregate a sequence of notes alternating rapidly between two pitch ranges into two perceptual streams (Bregman & Campbell, 1971; Dowling, 1973; McAdams & Bregman, 1979), so do infants (Demany, 1982). A converging result of Thorpe and Trehub (1989) illustrates this. Thorpe and Trehub played infants repeating six-note sequences such as AAAEEE (where A and E have frequencies of 440 and 660 Hz, a musical fifth apart). They trained the infants to turn their heads to see a toy whenever they heard a change in the stimuli being presented. A background pattern (AAAEEE) would be played over and over. Once in a while a changed pattern would appear. The changes consisted of temporal gaps introduced within perceptual groups (AAAE EE) or between groups (AAA EEE). The infants noticed the changes when they occurred within groups, but not between groups. An additional gap separating patterns that were already perceptually separate was simply lost in processing (as it tends to be by adults).

3. Pitch Infant pitch perception is quite accurate and also displays some of the sophistication of adult pitch processing. Adults display "octave equivalence" in being able to distinguish easily between a pair of tones an octave apart and a pair of tones not quite an octave apart (Ward, 1954), and so do infants (Demany & Armand, 1984). Adults also have "pitch constancy" in the sense that complex tones with differing harmonic structure (such as different vowel sounds with different frequency spectra) have the same pitch as long as their fundamental frequencies are the same. That is, we can sing "ah" and "ooh" on the same pitch, the listener will hear them that way, and the pitch can be varied

independently of vowel timbre by changing our vocal chord vibration rate (and hence the fundamental frequency of the vowel).

Even eliminating the fundamental frequency entirely from a complex tone will not change the pitch as long as several harmonics remain intact (Schouten, Ritsma, & Cardozo, 1962). Clarkson and Clifton (1985) used conditioned head turning to demonstrate that the same is true for infants 7 or 8 months old. Also, Clarkson and Rogers (1995) showed that, just like adults, infants have difficulty discerning the pitch when the harmonics that are present are high in frequency and remote from the frequency of the missing fundamental.

Regarding pitch discrimination, Thorpe (1986, as cited in Trehub, 1987) demonstrated that infants 7–10 months old can discriminate direction of pitch change for intervals as small as 1 semitone. Infants 6–9 months old can also be induced to match the pitches of vowels that are sung to them (Kessen, Levine, & Wendrich,1979; Révész, 1954; Shuter-Dyson & Gabriel, 1981).

4. Melodic Pitch Patterns Since early demonstrations by Melson and McCall (1970) and Kinney and Kagan (1976) that infants notice changes in melodies, a substantial body of research by Trehub (1985, 1987, 1990; Trehub & Trainor, 1990) and her colleagues has explored the importance for infants of a variety of dimensions of melodies. Figure 20.1 illustrates kinds of changes we can make in the pitch pattern of a melody, in this case "Twinkle, Twinkle, Little Star." We can shift the whole melody to a new pitch level, creating a transposition that leaves the pitch pattern in terms of exact intervals from note to note intact (fig-

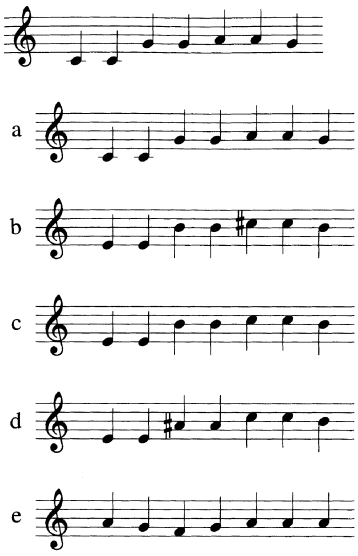


Figure 20.1
Examples of types of stimuli described in the text. At the top is the first phrase of the familiar melody, "Twinkle, Twinkle, Little Star," with the intervals between successive notes in semitones of [0, +7, 0, +2, 0, −2]. Following it are (a) an exact repetition of [0, +7, 0, +2, 0, −2]; (b) a transposition to another key [0, +7, 0, +2, 0, −2]; (c) a tonal imitation in the key of the original [0, +7, 0, +1, 0, −1]; (d) an imitation not in any major key [0, +6, 0, +2, 0, −1]; and (e) a melody with a different contour ("Mary Had a Little Lamb") [−2, −2, +2, +2, 0, 0].

ure 20.1b). We can shift the melody in pitch while preserving its contour (pattern of ups and downs) but changing its exact interval pattern (figures 20.1c and 20.1d), creating a same-contour imitation. The altered pitches of the same-contour imitation in figure 20.1c remain within a diatonic major scale, while those in figure 20.1d depart from it. Finally, we can change the contour (figure 20.1e), producing a completely different melody. Changes of contour are easily noticed by adults, whereas patterns with diatonic changes of intervals (figure 20.1c) are often hard to discriminate from transpositions (figure 20.1b; Dowling, 1978; Dowling & Fujitani, 1971).

Chang and Trehub (1977a) used heart-rate deceleration to indicate when a 5-month-old notices something new. Babies adapted to a continuously repeating six-note melody. Then Chang and Trehub substituted an altered melody to see if the baby would notice. When the stimulus was simply transposed 3 semitones (leaving it in much the same pitch range as before) the babies did not notice, but when the melody was shifted 3 semitones in pitch and its contour was altered, the babies showed a heart-rate deceleration "startle" response. For infants as for adults, the transposition sounds like the same old melody again, whereas the different-contour melody sounds new.

This result was refined in a study of 8- to 10-month-olds by Trehub, Bull, and Thorpe (1984). As in Thorpe and Trehub's (1989) study just described, Trehub et al. used conditioned head turning as an index of the infant's noticing changes in the melody. A background melody was played over and over. When a comparison melody replaced the background melody on a trial, the infants were able to notice all the changes Trehub et al. used: transpositions, same-contour-different-interval imitations, different-contour patterns, and patterns in which individual notes were displaced by an octave in a way that either violated, or did not violate, the contour. In this last transformation, the changes preserved *pitch class* by substituting a note an octave away that changed the contour. Pitch class depends on octave equivalence; all the members of a pitch class lie at octave multiples from each other. Contour changes were most noticeable. In a second experiment, Trehub et al. used the same task but made it more difficult by interposing three extra tones before the presentation of the comparison melody. In that case, infants did not notice the shift to transpositions and contour-preserving imitations, but they did notice changes in contour. This result was replicated with stimuli having even subtler contour changes by Trehub, Thorpe, and Morrongiello (1985).

The foregoing studies show that infants, like adults, easily notice differences in melodic contour. But, as Trehub, Thorpe, and Morrongiello (1987) point out, the studies do not demonstrate that infants in fact treat contour as a feature of melodies to be remembered. To show that, we would need to show that infants were abstracting a common property, an invariant, from a family of similar melodies that share only contour, and contrasting that property with that of melodies from another family with a different contour. To accomplish this, Trehub et al. (1987) used the conditioned-head-turning paradigm but with a series of background patterns that varied. In one condition, the background melodies varied in key and were all transpositions of one another. In a second condition, the background melodies were all contour-preserving imitations of one another, but not exact transpositions. In fact, infants were able to notice changes

among the background melodies, which were changes involving pitches (in the transposition set) and both intervals and pitches (in the imitation set). But they noticed changes of contour even more, supporting the notion that infants, like adults, encode and remember the contours of melodies they hear.

The results reviewed so far suggest considerable qualitative similarity between infants and adults in their memory for melodies. Both are able to notice changes in intervals and pitch levels of melodies under favorable conditions, but both find changes of melodic contour much more salient. The principal differences between infants and adults in the processing of pitch information in melodies arise from the acculturation of the adults in the tonal scale system of a particular culture. Virtually every culture in the world has at least one systematic pattern for the organization of pitch classes that repeats from octave to octave (Dowling & Harwood, 1986). The most common pattern in Western European music is that of the major ("*do, re, mi*") scale. Melodies that conform to that pattern are easier for Western European adults to encode and remember than melodies that do not (Cuddy, Cohen, & Mewhort, 1981; Dowling, 1991). However, as can be inferred from their cross-cultural variation, such scale patterns are not innate. There is no reason a priori for infants to find one pitch pattern easier than another.

This last point will probably strike psychologists as noncontroversial, but there is a very strong tradition among theorists of Western music going back to Pythagoras that attributes the structure of the Western scale system not only to innate cognitive tendencies, but, even further, to the structure of the universe itself in terms of simple whole-number ratios (Bernstein, 1976; Helmholtz, 1877/1954; Hindemith, 1961). The most sensible answer to these questions appears to be that there are certain constraints of human cognition that apply to musical scale structures but that within those constraints a very wide range of cultural variation occurs (Dowling & Harwood, 1986). The main constraints are octave equivalence (involving a 2/1 frequency ratio), a weaker tendency to give importance to the perfect fifth (a 3/2 ratio), coupled with a limit of seven or so pitch classes within the octave, in agreement with George Miller's (1956) argument concerning the number of categories along a perceptual dimension that humans can handle.

In a study bearing on the inherent importance of the perfect fifth, Trehub, Cohen, Thorpe, and Morrongiello (1986) used conditioned head turning to assess the performance of 9- to 11-month-olds in detecting changes of single pitches in a simple diatonic melody (C-E-G-E-C) and in a corresponding nondiatonic melody with an augmented fifth (C-E-G♯-E-C). They found no difference between the two background melodies, suggesting the lack of a strong inherent preference for the size of the fifth. Children between 4 and 6 years of age, however, did show a difference favoring the diatonic melody. Thus acculturation in the tonal scale system is already well begun by that age.

There is some evidence, however, in favor of the primacy of the perfect fifth. Cohen, Thorpe, and Trehub (1987) complicated the task used by Trehub et al. (1986) by transposing the background melody to a new pitch level with each repetition. In that case, the task could not be solved simply by noticing changes of single pitches, but would require the abstraction of the invariant interval

pattern of the background melody. Under those conditions, 7- to 11-month-olds found changes easier to detect in the diatonic pattern (C-E-G-E-C) than in the nondiatonic pattern (C-E-G♯-E-C). Seven to 11 months is a rather wide age range in the life of a rapidly changing infant. Lynch and Eilers (1992) differentiated the ends of that range by running 6-month-olds and 12-month-olds in parallel tasks. They found that although the 12-month-olds performed like the 7- to 11-month-olds in the Cohen, Thorpe, and Trehub (1987) study, the 6-month-olds performed equally well with the diatonic and nondiatonic patterns. That is, the younger infants were not yet acculturated to the standard Western diatonic scale as distinct from other arrangements of semitone intervals, whereas the older infants were.

In addition to the diatonic and nondiatonic patterns using Western "tonal material" (Dowling, 1978) consisting of intervals constructed of semitones, Lynch and Eilers (1992) also included a non-Western pattern: a Javanese *pélog* scale pattern that did not contain a perfect fifth and in which some of the pitches approximated quarter steps lying in between the semitones on the piano. The performance of the 6-month-olds, which was better than chance (and equally good) for diatonic and nondiatonic Western patterns, decreased to chance levels for the Javanese pattern (as did the performance of the 12-month-olds). Thus the 6-month-olds were either acculturated at the level of Western tonal material, or there is something about scale structures constructed with a logarithmic modulus such as the semitone (shared by the diatonic and nondiatonic patterns) that makes patterns constructed in them naturally easier to process. I favor the former explanation in terms of acculturation, because if conformity to "natural" pitch intervals were important, the most obvious candidate for a natural interval conducive to "good" pattern construction (in the Gestalt sense) is the perfect fifth (C-G, the 3/2 ratio) contained in the diatonic but not the other two patterns. This possibility is suggested by Trainor (1993), Trehub, Thorpe, and Trainor (1990), and Schellenberg and Trehub (1994) in their discussions of the diatonic/nondiatonic distinction made by the older infants. The perfect fifth is a fundamental building block in the traditional scale systems of India, China, and the American Indians, as well as of Europe (Dowling & Harwood, 1986), and is represented in the harmonic structure of complex tones such as vowel sounds, and also is prevalent in music (as at the start of "Twinkle, Twinkle," Figure 20.1). Thus if the perfect fifth, as a natural interval, were an important determinant of infant responses to scale patterns, the 6-month-olds would have performed better with the diatonic patterns than with the other two patterns. They did not, so it seems unlikely to me that the semitone, rarely explicitly present in the patterns and a far more remote candidate for natural interval, would play such a role.

If the younger infants are acculturated in terms of semitones, it remains nevertheless true that they are not sensitive to subtler aspects of the diatonic scheme. This is seen in their indifference both to the diatonic/nondiatonic distinction and to diatonic key membership of target tones, as shown by Trainor and Trehub (1992). Trainor and Trehub tested 8-month-olds using a strongly diatonic background melody. Comparison melodies had an altered pitch that either remained within the key of the background melody or went outside it. Infants

detected the change equally well whether it remained within the key or not. Their performance was unaffected by tonal scale structure. Adults, in contrast, found out-of-key alterations much easier to detect. (In fact, out-of-key alterations sound quite startling to adults unless they are "anchored" to a new key as the result of modulation—Bartlett, 1993; Bartlett & Dowling, 1988; Bharucha, 1984, 1996.) In fact, infants' performance with within-key alterations was superior to that of adults! Adults found the within-key alterations difficult to detect because the tonal framework they had acquired through lifelong perceptual learning made the within-key notes sound like natural continuations of the melody, even though they were the wrong notes. (Trainor & Trehub, 1993, extended these results to show that infants were more sensitive to changes in both patterns when they were transposed to a closely related key vs. a distant key—see the discussion of key-distance effects later.)

In summary, we can say that infants, like adults, find melodic contour a very salient feature of melodies. However, the process of acculturation in pitch-scale patterns is a long, slow process. By 6 months the infant is beginning that process at the level of the tonal material. By 1 year the infant responds differently to diatonic and nondiatonic patterns. But, as described below, listeners require more years of acculturation before they hear pitches automatically in terms of a tonal frame of reference.

5. Rhythm As noted in the earlier discussion of perceptual grouping, infants' temporal grouping of tone sequences is much like that of adults. Infants have been shown to discriminate between different rhythmic patterns (Chang & Trehub, 1977b; Demany, McKenzie, & Vurpillot, 1977). However, those tasks could have been solved on the basis of absolute rather than relative temporal relationships. Just as a melody retains its identity across transposition, so that relative and not absolute pitches are important, so a rhythmic pattern retains its identity across changes in tempo, where relative rather than absolute timing of the notes is important (Monahan & Carterette, 1985). And just as infants are sensitive to changes in patterns of relative pitch, they are sensitive to changes in the relative temporal patterns of rhythms. Trehub and Thorpe (1989), again using conditioned head turning, showed that infants 7–9 months old could notice changes in rhythmic patterns (such as XX XX vs. XXX X) even across variations in tempo. Just as for adults, a rhythmic pattern retained its identity when presented faster or slower.

Infants' broader rhythmic organization of musical phrases is like adults' in a surprising way. Krumhansl and Jusczyk (1990) presented 4- and 5-month-olds with Mozart minuets that had pauses inserted between phrases or within phrases. The infants preferred to listen to versions with pauses between phrases, suggesting that the infants were sensitive to cues to adult phrase structure of musical pieces. It remains to be seen exactly what cues the infants were responding to. Jusczyk and Krumhansl (1993) extended those results to show that the infants were really responding to phrase structure (and not just Mozart's beginning and ending patterns in the minuets) and that the pitch contour and note duration are important determinants of the infants' response to structural pauses. Furthermore, infants tended not to notice pauses inserted at phrase boundaries in naturally segmented minuets.

B. Childhood

During their second year, children begin to recognize certain melodies as stable entities in their environment and can identify them even after a considerable delay. My older daughter at 18 months would run to the TV set when she heard the "Sesame Street" theme come on, but not for other tunes. At 20 months, after a week or so of going around the house singing "uh-oh" rather loudly to a descending minor third, she responded with the spoken label "uh-oh" when I played that pattern on the piano.

1. Singing Children begin to sing spontaneously somewhere around the age of 9 months or a year. At first this can take the form of vocal play that includes wild excursions over the child's entire pitch range, but it also includes patterns of vowel sounds sung on locally stable pitches. This last is a feature that distinguishes singing from the child's incipient speech at this age.

Especially after 18 months, the child begins to generate recognizable, repeatable songs (Ostwald, 1973). The songs of a child around the age of 2 years often consist of brief phrases repeated over and over. Their contours are replicable, but the pitch wanders. The same melodic and rhythmic contour is repeated at different pitch levels, usually with different intervals between the notes. The rhythm of these phrases is coherent, with rhythms often those of speech patterns. Accents within phrases and the timing of the phrases themselves is determined by a regular beat pattern. This two-level organization of beat and within-phrase rhythm is another feature that distinguishes singing from speech and is characteristic of adult musical organization (Dowling, 1988; Dowling & Harwood, 1986).

An example of a spontaneous song from my daughter at 24 months consisted of an ascending and descending phrase with the words "Come a duck on my house" repeated 10 or 12 times at different pitch levels with small pitch intervals within phrases. This song recurred for 2 weeks and then disappeared. Such spontaneous songs have a systematic form and display two essential features of adult singing: they use discrete pitch levels, and they use the repetition of rhythmic and melodic contours as a formal device. They are unlike adult songs, however, because they lack a stable pitch framework (a scale) and use a very limited set of phrase contours in one song—usually just one or two (Dowling, 1984). A more sophisticated construction by the same child at 32 months can be seen in figure 20.2. The pitch still wanders but is locally stable within phrases. Here three identifiable phrases are built into a coherent song.

The preceding observations are in general agreement with those of Davidson, McKernon, and Gardner (1981; Davidson, 1985; McKernon, 1979) on spontaneous singing by 2-year-olds. Davidson et al. extended naturalistic observation by teaching a simple song to children across the preschool age range. Two- and 3-year-olds generally succeeded in reproducing the contours of isolated phrases. Older children were able to concatenate more phrases in closer approximations to the model. It was only very gradually across age that the interval relationships of the major scale began to stabilize. Four-year-olds could stick to a stable scale pattern within a phrase but would often slip to a new key for the next phrase, just as the 3-year-old in figure 20.2. It was not until after age 5 that the children could hold onto a stable tonality throughout the song. Further, with a

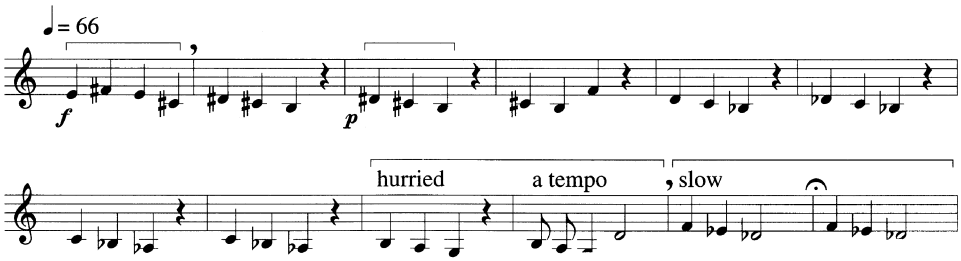


Figure 20.2
A child's spontaneous song at 32 months. Each note was vocalized to the syllable "Yeah." Brackets indicate regions of relatively accurate intonation. Elsewhere intonation wandered.

little practice, 5-year-olds were able to produce easily recognizable versions of the model. My own observations suggest that the typical 5-year-old has a fairly large repertoire of nursery songs of his or her culture. This emerges when children are asked to sing a song and can respond with a great variety of instances. It is also apparent from their better performance on memory tasks using familiar materials (vs. novel melodies; Andrews & Dowling, 1991). Through the preschool years, the use of more or less stable tonalities for songs comes to be established.

2. *Absolute Pitch* Absolute pitch is the ability to identify pitches by their note names even in the absence of musical context. Absolute pitch is not an essential ability for the understanding of most music, although it can aid in the tracking of key relationships in extended passages of tonal music (as in Mozart and Wagner) and in singing 12-tone music on sight. There are times when it can be a hindrance to music cognition by discouraging some of its possessors from developing sophisticated strategies for identifying pitch relationships in tonal contexts (Miyazaki, 1993). Absolute pitch has typically been quite rare even among musicians, occurring in only about 4–8%. However, in cultures where early music training is encouraged, such as in present-day Japan, the incidence of absolute pitch among the musically trained is much higher, possibly near 50% (Miyazaki, 1988). Ogawa and Miyazaki (1994) suggest on the basis of studies of 4- to 10-year-old children in a keyboard training program that most children have the underlying ability to acquire absolute pitch. In their review of the literature, Takeuchi and Hulse (1993) argue in favor of an "early-learning" hypothesis—that absolute pitch can be acquired by anyone, but only during a critical period ending in the fifth or sixth year.

Although relatively few adults can identify pitches, adults typically are able to approximate the pitch levels of familiar songs, a capacity that Takeuchi and Hulse (1993) call "residual absolute pitch." For example, Halpern (1989) found that adults would typically begin the same song on close to the same pitch after an extended delay. Levitin (1994), using the album cover as a retrieval cue, found that young adults sang popular songs they had heard only in one recorded version at approximately the correct pitch level. (Two thirds of the subjects were within 2 semitones of the correct pitch.)

The studies on pitch encoding cited earlier (Dowling, 1986, 1992) suggest that with a moderate amount of training people develop a "temporary and local" sense of absolute pitch that leads them to encode what they hear (and produce) in terms of the tonal framework provided by the current context.

3. Melodic Contour and Tonality In perception and in singing, melodic contour remains an important basis for melodic organization throughout childhood. Morrongiello, Trehub, Thorpe, and Capodilupo (1985) found 4- to 6-year-olds very capable in discriminating melodies on the basis of contour. Pick, Palmer, Hennessy, Unze, Jones, and Richardson (1988) replicated that result and found that 4- to 6-year-olds could also use contour to recognize same-contour imitations of familiar melodies. In another task emphasizing the recognition of similarity among same-contour imitations of familiar tunes, Andrews and Dowling (1991) found 5- and 6-year-olds performed equally well at recognizing familiar versions and both tonal and atonal imitations. It was not until ages 7 and 8 that tonality began to be a factor in that experiment and only by ages 9 or 10 that a difference appeared between familiar versions and same-contour imitations (the adult pattern of performance).

Studies of perception and memory provide converging evidence with that from singing concerning the 5- or 6-year-old's acquisition of a stable scale structure. With highly familiar tunes such as "Happy Birthday" and "Twinkle, Twinkle," even 4-year-olds can notice "funny" sounding versions with out-of-key pitches (Trehub, Morrongiello, & Thorpe, 1985). And Bartlett and Dowling (1980, Experiment 4) found that 5-year-olds can use musical key differences to discriminate between melodies. On each trial of the experiment, a familiar melody was presented, followed by either a transposition or a same-contour imitation. The comparison was either in the same key as the standard or a nearly related key, or it was in a distant key. (Near keys share many overlapping pitches in their scales; distant keys share few.) Adults in this task are highly accurate in saying "Same" to transpositions (>90%) and not saying "Same" to imitations (<10%). The pattern for 5-year-olds was very different: they tend to say "Same" to near-key comparisons (both transpositions and imitations) and "different" to far-key comparisons. Five-year-olds have one component of the adult behavior pattern—the ability to distinguish near from far keys—but not the other component—the ability to detect changes of interval sizes in the tonal imitations. They accept same-contour imitations as versions of the tune. As the child grows older, the pattern of response moves in the adult direction, so that an 8-year-old accepts near-key imitations less often than far-key transpositions. Eight-year-olds can use both key distance and interval changes to reject a same-contour imitation, whereas 5-year-olds rely principally on key distance.

The 5- to 6-year-old's grasp of stable tonal centers fits other results in the literature. For example, in a series of studies Riley and McKee (1963; Riley, McKee, Bell, & Schwartz, 1967; Riley, McKee & Hadley, 1964) found that first graders have an overwhelming tendency to respond by choosing a pitch match rather than an interval match. This tendency to respond to the pitch tasks in terms of a stable frame of reference contrasted with the same children's ability to respond to loudness-comparison tasks in terms of relative (not absolute) loudness.

The emergence of tonal scale relationships among the child's cognitive structures has implications for the conduct of research. Using atonal materials with infants has little impact on the results, because babies do not respond to tonal scale structures as such (Trainor & Trehub, 1992). But Wohlwill's (1971) use of atonal (and to the adult ear rather strange sounding) melodies probably led to his result that first graders could distinguish targets from different-contour lures at a level barely better than chance. At any rate, Wohlwill's conclusion that "the establishment of pitch as a directional dimension is a relatively late phenomenon" could not be true in the light of Thorpe's result with infants (1986, cited in Trehub, 1987). What is true is that first graders have trouble using words to describe pitch direction (Hair, 1977; Zimmerman & Sechrest, 1970).

During later childhood, the child continues to develop sophistication in the use of the tonal scale framework determined by the culture. This progress is illustrated by Zenatti (1969), who studied memory for sequences of three, four, and six notes with subjects from age 5 years up. On each trial, a standard melody was followed by a comparison melody in which one note of the standard had been changed by 1 or 2 semitones. The subject had to say which of the notes had been changed—a very difficult task. Zenatti found that for the three-note sequences, 5-year-olds performed at about chance with both tonal and atonal stimuli. From ages 6 through 10, the results for tonal and atonal sequences diverged, with better performance on tonal sequences. Then, at around age 12, processing of the atonal sequences caught up. For four- and six-note sequences, the same pattern appeared, but the tonal-atonal difference remained until adulthood. Experience with the tonal scale system leads people to improve on recognition of tonal melodies but not atonal melodies. With simple stimuli such as the three-note melodies, atonal performance catches up relatively soon, but longer sequences continue to benefit from the tonal framework throughout childhood. (This result converges with that of Morrongiello & Roes, 1990.) Superiority of recognition with tonal materials has been often observed with adults (Dowling, 1978; Francès, 1958/1988); Zenatti's study shows that the effect can be used as an index of the child's acquisition of the scale structures of the culture.

Trainor and Trehub (1994) took the development of the role of tonality in the ability to detect melodic pitch changes one step further. In addition to alterations that either remained within key or departed from the key, Trainor and Trehub introduced changes that remained in the key but departed from the particular harmony implied by the melody. For example, the first four notes of "Twinkle, Twinkle" (Figure la: C-C-G-G) imply harmonization with the tonic triad (C-E-G). A change of the third note from G to E would remain within both the key and the implied harmony. A change to F would remain within the key, but violate the harmony. Trainor and Trehub found that 7-year-olds, like adults, could detect the out-of-key and out-of-harmony changes much more easily than the within-harmony changes, whereas 5-year-olds reliably detected only the out-of-key changes. As Trainor and Trehub (1994, p. 131) conclude, "5-year-olds have implicit knowledge of key membership but not of implied harmony, whereas 7-year-olds, like adults, have implicit knowledge of both aspects of musical structure." In a result that converges with these studies, Imberty (1969, chapter 4) found that 7-year-olds could tell when a melody had

been switched in midstream from one key to another or from the major mode to the minor.

Krumhansl and Keil (1982) provide a good picture of the child's progress in grasping the tonal framework. They had children judge the goodness of melodic patterns beginning with an outline of the tonic triad (C-E-G) and ending on an arbitrarily chosen pitch. Krumhansl (1990) had found that adults in that task, especially musically experienced adults, produce a profile in which important notes in the tonal hierarchy (such as those of the tonic triad) receive high ratings and less important notes receive progressively lower ratings in accordance with their importance in the key. Krumhansl and Keil found that 6- and 7-year-olds distinguished simply between within-key notes and outside-of-key notes. The structure of the tonal hierarchy became more differentiated with age, so that by the age of 8 or 9 children were distinguishing between the pitches of the tonic triad and the other pitches within the key.

Two similar studies illustrate the importance of seemingly minor methodological details in research on the development of the tonal hierarchy. Cuddy and Badertscher (1987) simplified the task by using patterns with five notes instead of six. In that case, even 6- and 7-year-olds displayed the principal features of the adult hierarchy. And Speer and Meeks (1985) used an unstable context of the first seven notes of a C-major scale, ending on B or D (in contrast to the stable triad context in Krumhansl & Keil, 1982), to find that 8- and 11-year-olds perform very much like adults.

Lamont and Cross (1994) criticize the use of triads and scales as contexts in the foregoing three studies on two grounds. First, they suggest that these prototypical contexts, always the same throughout a condition of the experiment, are not very representative of the varied character of real tonal music. Second, they note that if children are exposed to any music class activities, the children will probably already have encountered scales and arpeggios. As Lamont and Cross (1994, p. 31) say, "Presented with an overlearned pattern,... the listener [could be expected] to give an overlearned response appropriate to that pattern." To produce more representative contexts, Lamont and Cross borrowed a method from West and Fryer (1990) of using a different random permutation of the notes of the major scale on each trial, and they also used chord progressions establishing the key. The study included five groups of children between 6 and 11 years old. Like Speer and Meeks (1985) and Cuddy and Badertscher (1987), Lamont and Cross found the children relatively sophisticated in their differentiation of the tonal hierarchy, but they also found, in agreement with Krumhansl and Keil (1982), that the children's representations of musical pitch gained in sophistication through the elementary school years. Lamont and Cross supplemented this study with converging evidence from a series of more open-ended tasks, such as arranging chime bars in order according to pitch and arranging them to create a tune.

In summary, the development of melody-processing skills can be seen as a progression from the use of gross, obvious features to the use of more and more subtle features. Babies can distinguish pitch contours and produce single pitches. Around the age of 5, the child can organize songs around stable tonal centers (keys) but does not yet have a stable tonal scale system that can be used to transpose melodies accurately to new keys. The scale system develops

during the elementary school years and confers on tonal materials an advantage in memory that remains into adulthood.

4. Rhythm There are two aspects of musical rhythm that I wish to discuss in terms of development in childhood. First is the development of the ability to control attention in relation to the temporal sequence of events, using regularities in the rhythm of occurrence of critical features in a piece to aim attention at important elements. Second is the development of the ability to remember and reproduce rhythmic patterns.

Adults in listening to speech and music are able to use their experience with similar patterns to focus their attention on critical moments in the ongoing stream of stimuli to pick up important information (Jones, 1981). This ability requires perceptual learning to develop. Andrews and Dowling (1991) studied the course of this development using a "hidden melodies" task in which the notes of a target melody such as "Twinkle, Twinkle" are temporally interleaved with random distractor notes in the same pitch range, the whole pattern being presented at 6 or 8 notes/sec. After about an hour of practice, adults can discern the hidden melody when they are told which target melody to listen for (Dowling, 1973; Dowling, Lung, & Herrbold, 1987). Andrews and Dowling (1991) included an easier condition in which the interleaved distractor notes were presented in a separate pitch range from the notes of the target. They reasoned that as listeners learned to aim attention in pitch, the listeners would find it easier to discern the targets in a separate pitch range. Five- and 6-year-olds perform barely better than chance on this task and find targets equally difficult to discern whether in a separate range from the distractors or not. It is not until the age of 9 or 10 that the separation of pitch ranges confers an advantage, suggesting that by that age listeners are able to aim their attention at a particular pitch range. Ability to aim attention in time improves steadily from age 6 on, and by age 9, discerning hidden targets with distractors in the same pitch range has reached 70% (with chance at 50%). Musically untrained adults achieve about 80% on this task, while musically experienced adults find the hidden targets equally easy to discern (about 90%) with distractors inside as well as outside the target pitch range.

There is evidence for the importance of a hierarchical organization of rhythm in 5-year-olds' reproductions of rhythmic patterns. Drake (1993) found 5-year-olds able to reproduce rhythms with two levels of organization: a steady beat and varying binary subdivisions of the beat. Although children that age find it easy to tap isochronous (steady, nonvarying) sequences in either binary or ternary rhythm, they find binary sequences with varying patterns within the beat easier than ternary. Drake reports that by the age of 7, children improve in reproducing models that include a variety of different durations in the same sequence, having gained facility with greater rhythmic complexity.

Accents in music can occur on various levels of structure. In particular, accents can be produced in terms of the two levels of beat and rhythmic organization. The beat or meter provides accents at regular time intervals. Rhythmic accents are generally conferred on the first and last members of rhythmic groups. A third level of accents can arise from discontinuities in the melodic contour, such as leaps and reversals of direction. Drake, Dowling, and Palmer

(1991) constructed songs in which accents on those levels either coincided or did not. Desynchronization of accent structure lowered children's performance in singing the songs, but there was little change in singing accuracy for children who are between 5 and 11 years old.

These results suggest that by the age of 5 children are responding to more than one level of rhythmic organization and that the songs they learn are processed as integrated wholes in the sense that events at one level affect performance at another; for example, complication of accent structure produces decrements in pitch accuracy in singing. An additional example is provided by Gérard and Auxiette (1988), who obtained rhythm reproductions from 5-year-olds. Gérard and Auxiette either provided the children with a plain rhythmic model to reproduce or provided additional context for the rhythm by providing either words to be chanted to it, or a melody to be sung to it, or both. They found that children with musical training performed best in tapping the rhythm when there was a melody, and children without musical training performed best when there were words. Having words or melody aided in the processing of the rhythm. Gérard and Auxiette (1992) also found that 6-year-old musicians were better able than nonmusicians to synchronize their tapping and their verbalizations in such a task.

The picture that emerges of the development of rhythmic organization is that a multilevel structure appears early and that by the age of 5, the child is quite sophisticated. There is some development in the school-age years, but Drake (1993), for example, found little difference between 7-year-olds and adult nonmusicians. Already the spontaneous songs of a 2-year-old show two levels of rhythmic organization, the beat and rhythmic subdivisions (often speech rhythms) overlaid on that, and the 5-year-old follows the same hierarchical organization in tapped reproductions. Finally, rhythmic organization is not easily separable from other aspects of structural organization in a song, so that in perception and production other aspects of melody are intertwined with rhythmic structure.

5. Emotion Ample evidence has accumulated that children during the preschool years learn to identify the emotional states represented in music, and this ability improves during the school years. For example, both Cunningham and Sterling (1988) and Dolgin and Adelson (1990) showed that by the age of 4, children perform well above chance in assigning one of four affective labels (essentially "happy," "sad," "angry," and "afraid") to musical excerpts in agreement with adults' choices. (With the exception of Cunningham and Sterling, all the studies reviewed here had subjects choose schematic faces expressing the emotions in making their responses.) Both of these studies also showed that performance improves over the school years. Performance was less than perfect at the earlier ages, and in particular, Cunningham and Sterling found that 4-year-olds were not consistently above chance with "sad" and "angry," nor 5-year-olds with "afraid," whereas Dolgin and Adelson found 4-year-olds at about chance with "afraid." In a similar study, Terwogt and Van Grinsven (1991) found that 5-year-olds performed very much like adults, but that all ages tended to confuse "afraid" and "angry." These studies were able in a general way to attribute the children's responses to features of the music, but there are

other studies that have focused on specific musical features such as the contrast between major and minor.

The issue of whether the major mode in Western music is a cue to happy emotions, and the minor mode a cue to sad ones, has been a perennial issue for both musicologists and psychologists. A particular developmental issue arises here, because we can ask whether responses to the affective connotations of major and minor appear earlier than the specific cognitive recognition of the difference, which, according to the foregoing review, appears around the age of 5. In exploring these issues, Gerardi and Gerken (1995) restricted responses to the choice of two faces, "happy" or "sad," and used adaptations of musical passages that differed in mode (major vs. minor) and predominant melodic contour (up vs. down). They found that 8-year-olds and adults, but not 5-year-olds, applied "happy" and "sad" consistently to excerpts in the major and minor, respectively. Only adults consistently chose "happy" for ascending contours and "sad" for descending, although that variable was probably not manipulated very strongly. (For example, "Che faro" from Gluck's *Orfeo ed Euridice* fails to ascend or descend unambiguously.)

In contrast to Gerardi and Gerken, Kastner and Crowder (1990) allowed subjects a choice of four faces—"happy," "neutral," "sad," and "angry"—and used versions of three different tunes presented in the major and minor, and with or without accompaniment. They found that when relatively positive responses (happy or neutral) were contrasted with negative responses (sad or angry), even 3-year-olds consistently assigned positive faces to major and negative faces to minor. This tendency became stronger between 3 and 12 years of age. Therefore, we can say that there is some indication that preschoolers are able to grasp the emotional connotations of the two modes at an earlier age than they can differentiate their responses in a more cognitively oriented task.

C. Adulthood

Rather than include here a comprehensive review of adults' implicit knowledge of musical structure, I shall concentrate on some issues concerned with tonality and the tonal scale framework. Adults in Western European cultures vary greatly in musical ability. Sometimes these individual differences are reflected in performance on perception and memory tasks. Untrained subjects usually do not find contour recognition more difficult than trained subjects (Dowling, 1978) but do find interval recognition (Bartlett & Dowling, 1980; Cuddy & Cohen, 1976) and the hearing out of partials in a complex tone (Fine & Moore, 1993) more difficult. Even where nonmusicians perform worse overall on tasks involving memory for melodies, they are often just as influenced as musicians by variables such as tonality, performing worse with atonal than with tonal melodies (Dowling, 1991). Also, nonmusicians are just as error prone as musicians when dealing with nonstandard quarter steps that fall in cracks in the musical scale (Dowling, 1992). Such qualitative results show that nonmusicians have acquired at least a basic tonal scale framework from their experience in the culture and that that framework has a psychological reality independent of its use as a pedagogical tool.

During the past few years, evidence has been accumulating that listeners routinely encode the music they hear in absolute, and not relative, terms. For

example, when presented with novel melodies and then tested after filled delays of up to 1.5 min, listeners find it easier to discriminate between targets (like figure 20.1b, only novel) and same-contour lures (like figure 20.1c), than between targets and different-contour lures (like figure 20.1e; Dowling, Kwak, & Andrews, 1995). (With familiar melodies such as those shown in figure 20.1, those abilities are about equal after 2 min.) That is, after a delay, listeners find it easier to discriminate very fine differences between the test melody and the melody they heard than to discriminate gross differences (DeWitt & Crowder, 1986; Dowling & Bartlett, 1981). Their memory represents very precisely what they have heard. This evidence converges with the demonstration by Levitin (1994), reviewed earlier, that nonmusicians come very close to the correct absolute pitch when singing familiar popular songs and with the similar demonstration by Levitin and Cook (1996) that their approximations of the tempos of such songs are quite accurate. This makes it seem likely that memory for music typically operates in terms of more precise representations of particular stimuli than has been generally thought (e.g., by Dowling, 1978).

Among adults, striking differences in performance based on different levels of musical experience sometimes appear, illustrating different ways in which knowledge of scale structure can be used. Dowling (1986) demonstrated differences among three levels of sophistication in a study of memory for novel seven-note melodies. Dowling presented the melodies in a context of chords that defined each melody as built around the tonic (the first degree of the scale, *do*) or the dominant (the fifth degree, *sol*). Listeners had to say whether notes had been altered when the melody was presented again. The test melodies were also presented with a chordal context, and that context was either the same as before or different. The test melodies were either exact transpositions or altered same-contour imitations of the original melodies. Musically untrained listeners performed equally well with same or different chord context at test. Listeners with moderate amounts of training in music (around 5 years of lessons when they were young) performed much worse with changed context. That suggests that those listeners were initially encoding the melodies in terms of the tonal scale values provided by the context, so that when the context was shifted, the melody was very difficult to retrieve. In contrast, nonmusicians simply remembered the melody independent of its relation to the context. Professional musicians performed very well with both changed and unchanged contexts. Their sophistication gave them the flexibility to ignore the context where it was not useful.

III. Summary

Adults bring a large store of implicit knowledge to bear in listening to music. This knowledge includes implicit representations of the tonal framework of the culture in terms of which expected events are processed efficiently and in terms of which pitches are interpreted in their musical context. This store of knowledge includes knowledge of the timing patterns of music in the culture, so that the listener is able to focus attention on moments in time at which critical information is likely to occur. Although musical experience leads, as we have seen, to greater sophistication in the store of implicit knowledge, nevertheless

nonmusicians have typically acquired the fundamentals of this knowledge from their experience listening to music throughout their lives. Thus nonmusicians are sensitive to shifts in tonality and to the multilevel structure of rhythmic organization.

The implicit knowledge of adults is built on elements present even in infancy: the importance of melodic and rhythmic contours, the use of discrete, steady pitch levels, the organization of rhythmic patterns into a steady beat and an overlay of more complicated rhythms, and octave equivalence, to name a few. These elements provide the groundwork for perceptual learning and acculturation throughout life to build upon.

Acknowledgment

I thank Melinda Andrews for her thoughtful contributions to the development of this chapter.

References

Andrews, M. W., & Dowling, W. J. (1991). The development of perception of interleaved melodies and control of auditory attention. *Music Perception, 8*, 349–368.

Bartlett, J. C. (1993). Tonal structure of melodies. In T. J. Tighe & W. J. Dowling (Eds.), *Psychology and music: The understanding of melody and rhythm* (pp. 39–61). Hillsdale, NJ: Erlbaum.

Bartlett, J. C., & Dowling, W. J. (1980). The recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception & Performance, 6*, 501–515.

Bartlett, J. C., & Dowling, W. J. (1988). Scale structure and similarity of melodies. *Music Perception, 5*, 285–314.

Bernstein, L. (1976). *The unanswered question*. Cambridge, MA: Harvard University Press.

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology, 16*, 485–518.

Bharucha, J. J. (1996). Melodic anchoring. *Music Perception, 13*, 383–400.

Bregman, A., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89*, 244–249.

Brown, R. (1973). *A first language: The early stages*. London: George Allen & Unwin.

Chang, H. W., & Trehub, S. E. (1977a). Auditory processing of relational information by young infants. *Journal of Experimental Child Psychology, 24*, 324–331.

Chang, H. W., & Trehub, S. E. (1977b). Infant's perception of temporal grouping in auditory patterns. *Child Development, 48*, 1666–1670.

Clarkson, M. G., & Clifton, R. K. (1985). Infant pitch perception: Evidence for responding to pitch categories and the missing fundamental. *Journal of the Acoustical Society of America, 77*, 1521–1528.

Clarkson, M. G., & Rogers, E. C. (1995). Infants require low-frequency energy to hear the pitch of the missing fundamental. *Journal of the Acoustical Society of America, 98*, 148–154.

Cohen, A. J., Thorpe, L. A., & Trehub, S. E. (1987). Infants' perception of musical relations in short transposed tone sequences. *Canadian Journal of Psychology, 41*, 33–47.

Cuddy, L. L., & Badertscher, B. (1987). Recovery of the tonal hierarchy: Some comparisons across age and levels of musical experience. *Perception & Psychophysics, 41*, 609–620.

Cuddy, L. L., & Cohen, A. J. (1976). Recognition of transposed melodic sequences. *Quarterly of Experimental Psychology, 28*, 255–270.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception & Performance, 7*, 869–883.

Cunningham, J. G., & Sterling, R. S. (1988). Developmental change in the understanding of affective meaning of music. *Motivation & Emotion, 12*, 399–413.

Davidson, L. (1985). Tonal structures in children's early songs. *Music Perception, 2*, 361–374.

Davidson, L., McKernon, P., & Gardner, H. (1981). The acquisition of song: A developmental approach. In *Documentary report of the Ann Arbor Symposium* (pp. 301–315). Reston, VA: Music Educators National Conference.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 208*, 1174–1176.

DeCasper, A. J., & Spence, M. J. (1986). Prematernal speech influences newborns' perception of speech sounds. *Infant Behavior & Development, 9*, 133–l50.

Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behavior & Development, 5*, 261–276.

Demany, L., & Armand, F. (1984). The perceptual reality of tone chroma in early infancy. *Journal of the Acoustical Society of America, 76*, 57–66.

Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature, 266*, 718–719.

DeWitt, L. A., & Crowder, R. G. (1986). Recognition of novel melodies after brief delays. *Music Perception, 3*, 259–274.

Dolgin, K. G., & Adelson, E. H. (1990). Age changes in the ability to interpret affect in sung and instrumentally-presented melodies. *Psychology of Music, 18*, 87–98.

Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology, 5*, 322–337.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85*, 341–354.

Dowling, W. J. (1984). Development of musical schemata in children's spontaneous singing. In W. R. Crozier & A. J. Chapman (Eds.), *Cognitive processes in the perception of art* (pp. 145–163). Amsterdam: North-Holland.

Dowling, W. J. (1986). Context effects on melody recognition: Scale-step versus interval representations. *Music Perception, 3*, 281–296.

Dowling, W. J. (1988). Tonal structure and children's early learning of music. In J. Sloboda (Ed.), *Generative processes in music* (pp. 113–128). Oxford: Oxford University Press.

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics, 50*, 305–313.

Dowling, W. J. (1992). Perceptual grouping, attention and expectancy in listening to music. In J. Sundberg (Ed.), *Gluing tones: Grouping in music composition, performance and listening* (pp. 77–98). Stockholm: Publications of the Royal Swedish Academy of Music, no. 72.

Dowling, W. J. (1993a). Procedural and declarative knowledge in music cognition and education. In T. J. Tighe & W. J. Dowling (Eds.), *Psychology and music: The understanding of melody and rhythm* (pp. 5–18). Hillsdale, NJ: Erlbaum.

Dowling, W. J. (1993b). La structuration melodique: Perception et chant. In A. Zenatti (Ed.), *Psychologie de la musique* (pp. 145–176). Paris: Presses Universitaires de France.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology, 1*(1), 30–49.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49*, 524–531.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. New York: Academic Press.

Dowling, W. J., Kwak, S.-Y., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics, 57*, 136–149.

Dowling, W. J., Lung, K. M.-T., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception & Psychophysics, 41*, 642–656.

Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics, 53*, 25–33.

Drake, C., Dowling, W. J., & Palmer, C. (1991). Accent structures in the reproduction of simple tunes by children and adult pianists. *Music Perception, 8*, 315–334.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science, 270*, 305–307.

Fine, P. A., & Moore, B. J. C. (1993). Frequency analysis and musical ability. *Music Perception, 11*, 39–54.

Francès, R. (1988). *The perception of music* (W. J. Dowling, Trans.). Hillsdale, NJ: Erlbaum. (Original publication 1958).

Gérard, C., & Auxiette, C. (1988). The role of melodic and verbal organization in the reproduction of rhythmic groups by children. *Music Perception, 6*, 173–192.

Gérard, C., & Auxiette, C. (1992). The processing of musical prosody by musical and nonmusical children. *Music Perception, 10*, 93–126.

Gerardi, G. M., & Gerken, L. (1995). The development of affective responses to modality and melodic contour. *Music Perception, 12*, 279–290.

Hair, H. I. (1977). Discrimination of tonal direction on verbal and nonverbal tasks by first-grade children. *Journal of Research on Music Education, 25*, 197–210.

Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition, 17*, 572–581.

Helmholtz, H. von. (1954). *On the sensations of tone*. (A. J. Ellis, Trans.). New York: Dover. (Original work published 1877.)

Hindemith, P. A. (1961). *Composer's world*. New York: Doubleday.

Imberty, M. (1969). *L'acquisition des structures tonales chez l'enfant*. Paris: Klincksieck.

Jones, M. R. (1981). Only time can tell: On the topology of mental space and time. *Critical Inquiry, 7*, 557–576.

Jusczyk, P. W., & Krumhansl, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception & Performance, 19*, 627–640.

Kastner, M. P., & Crowder, R. G. (1990). Perception of major/minor: IV. Emotional connotations in young children. *Music Perception, 8*, 189–202.

Kessen, W., Levine, J., & Wendrich, K. A. (1979). The imitation of pitch in infants. *Infant Behavior & Development, 2*, 93–99.

Kinney, D. K., & Kagan, J. (1976). Infant attention to auditory discrepancy. *Child Development, 47*, 155–164.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science, 1*, 70–73.

Krumhansl, C. L., & Keil, F. C. (1982). Acquisition of the hierarchy of tonal functions in music. *Memory & Cognition, 10*, 243–251.

Lamont, A., & Cross, I. (1994). Children's cognitive representations of musical pitch. *Music Perception, 12*, 27–55.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics, 56*, 414–423.

Levitin, D. J., & Cook, P. R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics, 58*, 927–935.

Lynch, M. P., & Eilers, R. E. (1992). A study of perceptual development for musical tuning. *Perception & Psychophysics, 52*, 599–608.

McAdams, S., & Bregman, A. (1979). Hearing musical streams. *Computer Music Journal, 3*(4), 26–43, 60.

McKernon, P. E. (1979). The development of first songs in young children. *New Directions for Child Development, 3*, 43–58.

Mehler, J., Bertoncini, J., Barrière, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception, 7*, 491–497.

Melson, W. H., & McCall, R. B. (1970). Attentional responses of five-month girls to discrepant auditory stimuli. *Child Development, 41*, 1159–1171.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Miyazaki, K. (1988). Musical pitch identification by absolute pitch possessors. *Perception & Psychophysics, 44*, 501–512.

Miyazaki, K. (1993). Absolute pitch as an inability: Identification of musical intervals in a tonal context. *Music Perception, 11*, 55–72.

Monahan, C. B., & Carterette, E. C. (1985). Pitch and duration as determinants of musical space. *Music Perception, 3*, 1–32.

Morrongiello, B. A., & Roes, C. L. (1990). Developmental changes in children's perception of musical sequences: Effects of musical training. *Developmental Psychology, 26*, 814–820.

Morrongiello, B. A., Trehub, S. E., Thorpe, L. A., & Capodilupo, S. (1985). Children's perception of melodies: The role of contour, frequency, and rate of presentation. *Journal of Experimental Child Psychology, 40*, 279–292.

Ogawa, Y., & Miyazaki, K. (1994, July). *The process of acquisition of absolute pitch by children in Yamaha music school*. Paper presented at the Third International Conference for Music Perception and Cognition, Liège, Belgium.

Ostwald, P. F. (1973). Musical behavior in early childhood. *Developmental Medicine & Child Neurology, 15*, 367–375.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*, 811.

Pick, A. D., Palmer, C. F., Hennessy, B. L., Unze, M. G., Jones, R. K., & Richardson, R. M. (1988). Children's perception of certain musical properties: Scale and contour. *Journal of Experimental Child Psychology, 45*, 28–51.

Révész, G. (1954). *Introduction to the psychology of music*. Norman: University of Oklahoma Press.

Riley, D. A., & McKee, J. P. (1963). Pitch and loudness transposition in children and adults. *Child Development, 34*, 471–483.

Riley, D. A., McKee, J. P., Bell, D. D., & Schwartz, C. R. (1967). Auditory discrimination in children: The effect of relative and absolute instructions on retention and transfer. *Journal of Experimental Psychology, 73*, 581–588.

Riley, D. A., McKee, J. P., & Hadley, R. W. (1964). Prediction of auditory discrimination learning and transposition from children's auditory ordering ability. *Journal of Experimental Psychology, 67*, 324–329.

Schellenberg, E. G., & Trehub, S. E. (1994). Frequency ratios and the perception of tone patterns. *Psychonomic Bulletin & Review, 2*, 191–201.

Schouten, J. F., Ritsma, B. J., & Cardozo, B. L. (1962). Pitch of the residue. *Journal of the Acoustical Society of America, 34*, 1418–1424.

Shetler, D. J. (1989). The inquiry into prenatal musical experience: A report of the Eastman Project, 1980–1987. *Pre- and Peri-Natal Psychology, 3*, 171–189.

Shuter-Dyson, R., & Gabriel, C. (1981). *The psychology of musical ability*. London: Methuen.

Speer, J. R., & Meeks, P. U. (1985). School children's perception of pitch in music. *Psychomusicology, 5*, 49–56.

Spence, M. J., & DeCasper, A. J. (1987). Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. *Infant Behavior & Development, 10*, 133–142.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin, 113*, 345–361.

Terwogt, M. M., & Van Grinsven, F. (1991). Musical expression of moodstates. *Psychology of Music, 19*, 99–109.

Thorpe, L. A., & Trehub, S. E. (1989). Duration illusion and auditory grouping in infancy. *Developmental Psychology, 25*, 122–127.

Trainor, L. J. (1993, March). *What makes a melody intrinsically easy to process: Comparing infant and adult listeners*. Paper presented to the Society of Research in Child Development, New Orleans.

Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western tonal structure. *Journal of Experimental Psychology: Human Perception & Performance, 18*, 394–402.

Trainor, L. J., & Trehub, S. E. (1993). Musical context effects in infants and adults: Key distance. *Journal of Experimental Psychology: Human Perception & Performance, 19*, 615–626.

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics, 56*, 125–132.

Trehub, S. E. (1985). Auditory pattern perception in infancy. In S. E. Trehub & B. A. Schneider (Eds.), Auditory development in infancy (pp. 183–l95). New York: Plenum.

Trehub, S. E. (1987). Infants' perception of musical patterns. *Perception & Psychophysics, 41*, 635–641.

Trehub, S. E. (1990). Human infants' perception of auditory patterns. *International Journal of Comparative Psychology, 4*, 91–110.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development, 55*, 821–830.

Trehub, S. E., Cohen, A. J., Thorpe, L. A., & Morrongiello, B. A. (1986). Development of the perception of musical relations: Semitone and diatonic structure. *Journal of Experimental Psychology: Human Perception & Performance, 12*, 295–301.

Trehub, S. E., Morrongiello, B. A., & Thorpe, L. A. (1985). Children's perception of familiar melodies: The role of intervals. *Psychomusicology, 5*, 39–48.

Trehub, S. E., & Thorpe, L. A. (1989). Infants' perception of rhythm: Categorization of auditory sequences by temporal structure. *Canadian Journal of Psychology, 43*, 217–229.

Trehub, S. E., Thorpe, L. A., & Morrongiello, B. A. (1985). Infants' perception of melodies: Changes in a single tone. *Infant Behavior & Development, 8*, 213–223.

Trehub, S. E., Thorpe, L. A., & Morrongiello, B. A. (1987). Organizational processes in infants' perception of auditory patterns. *Child Development, 58*, 741–749.

Trehub, S. E., Thorpe, L. A., & Trainor, L. J. (1990). Infants' perception of good and bad melodies. *Psychomusicology, 9*, 5–19.

Trehub, S. E., & Trainor, L. J. (1990). Rules for listening in infancy. In J. Enns (Ed.), *The development of attention: Research and theory* (pp. 87–119). Amsterdam: Elsevier.

Ward, W. D. (1954). Subjective musical pitch. *Journal of the Acoustical Society of America, 26*, 369–380.

West, R. J., & Fryer, R. (1990). Ratings of suitability of probe tones as tonics after random ordering of notes of the diatonic scale. *Music Perception, 7*, 253–258.

Wohlwill, J. F. (1971). Effect of correlated visual and tactual feedback on auditory pattern learning at different age levels. *Journal of Experimental Child Psychology, 11*, 213–228.

Zenatti, A. (1969). Le développement génétique de la perception musicale. *Monographies Françaises de Psychologie, No. 17*.

Zimmerman, M. P., & Sechrest, L. (1970). Brief focused instruction and musical concepts. *Journal of Research on Music Education, 18*, 25–36.

Chapter 21

Cognitive Psychology and Music

Roger N. Shepard and Daniel J. Levitin

21.1 Cognitive Psychology

What does cognitive psychology have to do with the perception of sound and music? There is a long chain of processes between the physical events going on in the world and the perceptual registration of those events by a human observer. The processes include the generation of energy by some external object or event, the transmission of the energy through the space between the event and the observer, the reception and processing of the energy by the observer's sensory receptors, and the transmission of signals to the brain, where still more processing takes place. Presumably, the end result is the formation of a representation in the brain of what is going on in the external world. The brain has been shaped by natural selection; only those organisms that were able to interpret correctly what goes on in the external world and to behave accordingly have survived to reproduce.

The way we experience all events in the world, including musical events, is the result of this process of interpretation in the brain. What is happening inside the eye on the surface of the retina, or on the basilar membrane in the ear, is of no significant interest whatsoever, except insofar as it provides information from which the brain is able to construct a representation of what is going on in the world. True, the signals from the receptors are generally the only source of information the brain has about what is actually going on in the external world, so it is important to understand the workings of the observer's eyes and ears. But what goes on in those sensory transducers has relatively little direct correspondence to the final representation experienced by the observer, which is the result of extensive further processing within the observer's brain.

Sensory psychophysicists and psychologists study what goes on in the sensory transducers, and the eye and ear appear fundamentally quite different in function and behavior. There are many things specific to a particular sensory organ, and they must be studied and discussed independently. In contrast, cognitive psychologists are principally interested in the final internal representation. If the internal representation is to be useful, it must correspond to events in the real world. There is one world to be perceived, and all of the senses provide information to the observer about that world. Therefore, a confluence should emerge from the processing in the brain, regardless of whether the

From chapter 3 in *Music, Cognition, and Computerized Sound*, ed. P. R. Cook (Cambridge, MA: MIT Press, 1999), 21–35. Reprinted with permission.

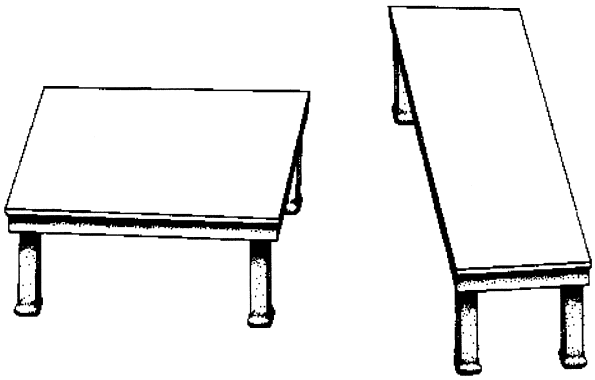


Figure 21.1
Things are sometimes different than they appear to be. Use a ruler to measure the tops of these two tables, specifically comparing the short ends and the long ends.

input is from the visual, auditory, or some other sensory modality. This chapter will point out some general principles of perception and cognition that, though similar for vision and audition, are directly relevant to the understanding of music and music perception.

Figure 21.1 demonstrates that internal representation can indeed be quite different from the physical stimulus on the retina. Two tables are depicted as if in different orientations in space, but stating that there are two tables already makes a cognitive interpretation. The figure actually consists only of a pattern of lines (or dots) on a two-dimensional surface. Still, humans tend to interpret the patterns of lines as three-dimensional objects, as two differently oriented tables with one larger than the other. If one were able to turn off the cognitive representation of "tables in space," one would see that the two parallelograms corresponding to the tabletops are of identical size and shape! Verify this with a ruler, or trace one parallelogram (tabletop) on a sheet of tracing paper and then slide it into congruence with the other. The fact that it is difficult to see the two tabletops abstractly as simple parallelograms, and thus to see them as the same size and shape, proves that the internal representation in the brain is quite different from the pattern present on the sensory surface (retina). We tend to represent the pattern of lines as objects in the external world because evolution has selected for such representation. The interpretation process in the brain has been shaped to be so automatic, swift, and efficient that it is virtually unconscious and outside of our control. As a result, we cannot suppress it even when we try.

21.2 Unconscious Inference

Hermann von Helmholtz (born 1821) made more contributions to the understanding of hearing and vision than perhaps any other individual. In addition to his fundamental contributions to physics and to physiology, in cognitive psychology he is known for his formulation of the principle of *unconscious inference*. Figure 21.2 illustrates the principle of unconscious inference. Our

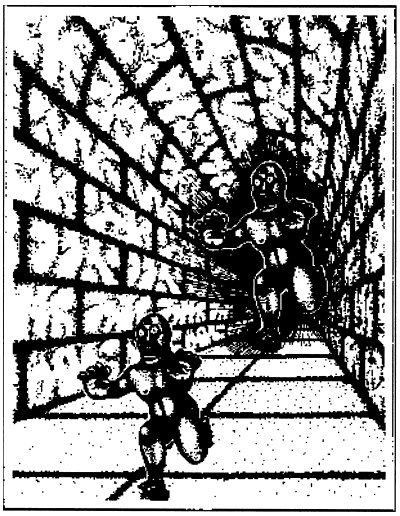


Figure 21.2
Unconscious inference is at work in this picture. Even though both "monsters" are exactly the same size (measure them with a ruler), the perspective placement makes the chaser look bigger than the one being chased.

perceptual machinery automatically makes the inference to three-dimensional objects on the basis of perceptual cues that are present in the two-dimensional pattern on the retina. Cues—particularly linear perspective—support the inference to the three-dimensional interpretation, but the inference is quite unconscious.

Many retinal cues enable us to construct a three-dimensional representation from purely two-dimensional representation input. Following are a few examples of these cues:

Linear perspective. Converging lines in a two-dimensional drawing convey parallel lines and depth in three dimensions. This is evident in the rows of stones in figure 21.2.

Gradient of size. The elements of a uniform texture decrease in size as they approach the horizon. This is evident in figure 21.2, where the stone patterns get smaller in the receding tunnel.

Aerial perspective. Objects in the far distance appear lighter and blue (for the same reason that the sky appears light and blue).

Binocular parallax. Each of our two eyes receives a slightly different image, and from these the brain is able to make quite precise inferences about the relative distances of objects. This is particularly true for objects close to the observer.

Motion parallax. Movement on the part of the observer changes the images on each retina, and the differences between successive viewpoints is used to infer distances, just as in binocular parallax.

It is interesting to note that in general we have no notion of the cues that our brains are using. Experiments have shown that some of the cues can be missing

(or intentionally removed); but as long as some subset of these cues is still available, the observer sees things in depth and can make accurate judgments about the relative distances and placements of objects. Even though the examples printed in this book are just two-dimensional drawings, the important thing to remember is that all images end up entering our retinas as two-dimensional images. We use unconscious inference to make sense of the real world just as we use it to interpret drawings, photographs, and movies.

The use of the term *inference* does not imply that the cognitive processes of interpretation are mere probabilistic guesses, although situations do occur in which the number of cues is reduced to the point where unconscious inference may become a random guesslike process. James Gibson, a perceptual psychologist at Cornell University, emphasized that under most circumstances (when there is good illumination, we are free to move about with both eyes open, and our spatial perception is completely accurate and certain), the information is sufficient to construct an accurate representation of the disposition of objects in space. Gibson referred to this as *direct perception*, as contrasted to *unconscious inference*. The two can be reconciled by the fact that complex computation must go on to process the information coming into the sensory systems, and most of that computation goes on unconsciously. The information is integrated in order to give very precise information about what is going on in the world, not random guesses based on fragmentary information.

21.3 Size and Loudness Constancy

Objects in the world are, in general, of constant size; but the image of an object on the retina expands and contracts as the object moves closer and farther away. What has been important for us and for our ancestors has been the ability to perceive objects as they are, independent of their distance from us. This is known as *size constancy*. Figure 21.3 demonstrates this principle.

In the auditory domain, *loudness constancy* is a direct analog of size constancy. If an instrument emitting a sound of constant output is moved farther away, the intensity that reaches a listener decreases. This is because the wave fronts emanating from the instrument are spherical in shape, and the surface area of a sphere increases with the square of the radius. The energy from the instrument is uniformly distributed over this spherical surface, and hence the intensity reaching the listener decreases with the square of the distance from the instrument to the listener. Not surprisingly then, if the amplitude of a sound is decreased, the sound may seem to come from farther away. But we could alternatively experience the source as decreasing in intensity without moving farther away. Similarly, a visually perceived balloon from which air is escaping may appear to be receding into the distance or simply shrinking in size. Other cues besides size or loudness may determine whether the change in the external world is in the size or the intensity of the source, or in its distance from the observer.

The intensity of a musical source can be decreased by playing the instrument more softly. There are accompanying changes in timbre, however, that are different from a simple decrease in amplitude. The higher-frequency components of the sound tend to increase and decrease with the effort exerted by the musi-

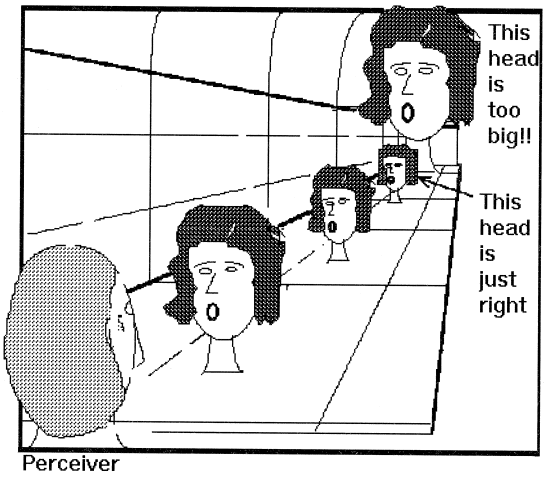


Figure 21.3
Size constancy. The head closest to the perceiver is the same physical size on the page as the "too-big" head farthest from the perceiver.

cian, an amount that is not proportional to the lower components of the spectrum. Thus, spectral balance as well as overall amplitude provides cues to the intensity versus distance of a source.

In our normal surroundings, there are surfaces around us that reflect sound, causing echoes or reverberation. In general we have little direct awareness of the reflected sound reaching us via these paths, but we use the information in these reflected waves to make unconscious inferences about the surroundings and sound sources within those surroundings. The reflections tell us, for example, that we are in a room of a certain size and composition, and give us a sense of the space. We receive a signal from a sound source within the room, then some time later we receive signals via the reflected paths. If a sound source is close, the direct sound is relatively intense, and the reflected sounds occur at decreased intensity and later in time. If the sound source moves away, the direct sound decreases, but the reflected sound remains roughly constant in intensity. The time difference between the arrival of the direct and the reflected sounds also decreases as the source recedes. By unconscious inference, the intensity ratio of direct to reflected sound, and the time delay between the direct and reflected sound, are used, along with other cues, to determine the distance and intensity of a source.

21.4 Spatial and Temporal Inversion

Some of the correlations in the world are so common that we have developed special machinery for their interpretation. If a familiar pattern is transformed in some way, even though all of the information is retained intact, then that pattern will not be interpreted in the same way by a human observer because our machinery is "wired" to interpret the information only in its usually encountered form. Consider the simple transformation of rotation. Figure 21.4 shows a

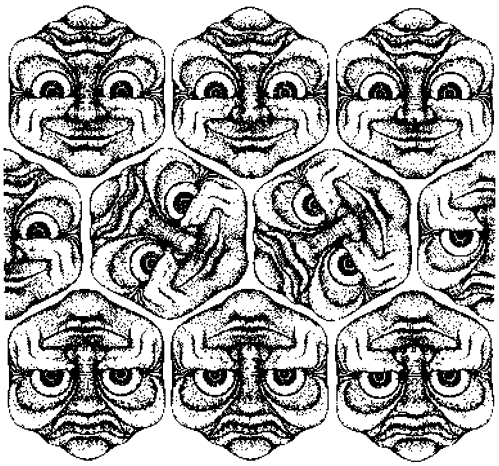


Figure 21.4
Turn this page over and you will still see faces right-side up. After infancy, we become more tuned to seeing faces right-side up, and thus must try hard to see the "frowning" faces as being upside down.

number of presentations of the same face. Because we are attuned to perceiving faces in their usual upright orientation, the upper and lower rows of shapes shown in the figure are perceived as being of two different faces rather than as one face in two orientations. We tend to make the interpretation that is consistent with a standard face, in which the eyes are on the top and the mouth is on the bottom. Developmental studies have shown that up to a certain age, children are equally skilled at interpreting faces either right-side up or upside down, but with increasing age the skill at interpreting faces right-side up continues to increase after the ability to interpret inverted faces levels off. Eventually the right-side up exposure becomes so great that the perception dominates. We develop an impressive ability to recognize and to interpret the expressions of right-side up faces—an ability not yet matched by machine—but this ability does not generalize to upside-down faces, with which we have had much less practice.

An analog of this spatial inversion in the visual domain is a temporal reversal in the auditory domain. In normal surroundings, we receive direct and reflected sound. We generally do not hear the reflected echoes and reverberation as such, but make the unconscious inference that we are hearing the source in a certain type of space, where the impression of that space is determined by the character of the reflected signals.

It is curious that the addition of walls and boundaries, essentially limiting space, gives the sense of spaciousness in audition. In a purely anechoic room (a specially constructed space that minimizes reflections from the walls, floor, and ceiling) we get no reverberation, and thus no sense of space. In vision, too, if an observer were in space with no objects around, there would be no sense of the space. Gibson pointed out that we do not perceive space but, rather, objects in space. In audition, we need surfaces to give us the sense of the space they define.

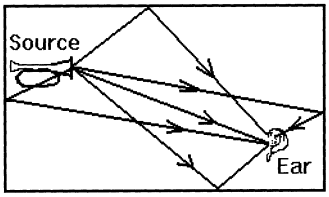


Figure 21.5
Many reflected acoustic paths in a room.

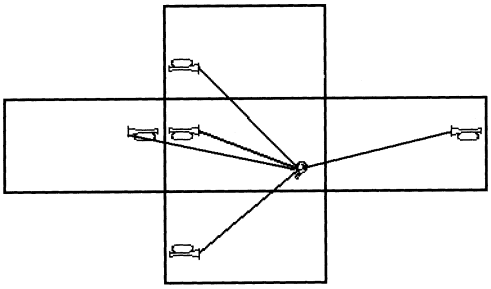


Figure 21.6
The same paths, shown as direct paths from "virtual sources."

The ears can hear the direction of the source by comparing the differences between the arrival times and intensities at the two ears. The ears can similarly process differences in times and amplitudes of reflected sounds, and infer the source locations implied by those reflected sounds. In this way, we auditorally identify a sound source and a number of *virtual sources*, or copies of the sound source in virtual locations that lie *outside* the space actually enclosed by the walls. Figure 21.5 shows a sound source, an ear, and a few reflected sound paths. Only the first reflections (those that reflect from only one wall in going from the source to the listener) are shown, but there are many important second, third, and so on reflections. Figure 21.6 shows the same sound paths as direct paths from virtual sources. It is clear why reverberation gives the sense of space, with virtual sources distributed over a large space outside the room. The same sense of space can be experienced visually in a room (such as a barbershop or restaurant) with large mirrors on opposite walls.

There is a fundamental time asymmetry in the reception of direct and reflected sounds. All reflected sounds reach the listener *after* the direct signal. This is a manifestation of the second law of thermodynamics: in the absence of external energy input, order tends to go over into disorder. The direct sound may be orderly, but the randomly timed reflected copies of that sound appear to become random, with a momentary impulse decaying into white noise over time. Our auditory processing machinery evolved to process echoes and reverberation that follow a direct signal; it is ill-equipped to deal with an artificially produced case in which the echoes precede the signal.

Similarly, a resonant object, when struck, typically produces a sound that decays exponentially with time. Because of its unfamiliarity, a note or chord

struck on a piano (with the damper lifted) sounds quite odd when played backward on a tape recorder. The sound suggests an organ more than a piano, slowly building up to an abrupt termination that gives no percussive impression. Moreover, if the individual notes of the chord are struck in rapid succession rather than simultaneously, their order is much more difficult to determine in the temporally reversed case than in the normal, forward presentation.

A sound of two hands clapping in a room is quite natural. In a normal-size room, the listener will hardly notice the reverberation; but in larger rooms it becomes noticeable, and the listener may think of the sounds only as indicating a large room, not a long reverberation. Completely unnatural sounds be can be created by mechanical manipulations using tape or a digital computer. A sound can be reversed, then reverberation added, then the resultant sound can be reversed. This generates a sound where the reverberation precedes the sound, but the sound itself still progresses forward. Speech processed in this fashion becomes extremely difficult to understand. This is because we are used to processing speech sounds in reverberant environments but are completely unfamiliar with an environment that would cause reverberation to come before a sound.

21.5 Perceptual Completion

Another fundamental principle of perception is called *perceptual completion*. Sometimes we have incomplete information coming into our sensory systems. To infer what is going on, we have to do some amount of top-down processing in addition to the normal bottom-up processing. We must complete the information to determine the most probable explanation for what is occurring in the real world that is consistent with the information presented to our senses. All of us can think of familiar examples of this from our own experience with camouflage, both in nature, with animals, insects, and birds, and in the artificial camouflage worn by humans. There are also many examples from the art world of the intentional use and manipulation of ambiguity and camouflage. Most famous perhaps are paintings by Bev Doolittle, such as her "Pintos on a Snowy Background," which depicts pinto horses against a snowy and rocky mountainside. Because of the patterning of the brown and white horse hair on the pintos, it is not easily distinguished from the background of brown rocks and white snow.

It is difficult to program a computer to correctly process ambiguous visual stimuli, because computers do not have the kind of real-world knowledge that humans have gained through evolution and learning. This knowledge allows us to make reasonable inferences about what is going on in the world, using only partial information. Figure 21.7 shows two (or more) objects, with one of the objects apparently covering part of the other object. The most probable explanation for the alignment of the objects is that the bar is one object that extends continuously under the disk. It is also possible that there are two shorter bars whose colors, alignments, and such just happen to coincide as shown in figure 21.8. But the simpler explanation is that it is a single bar. Research with young infants has shown that they, too, are sensitive to this type of environ-

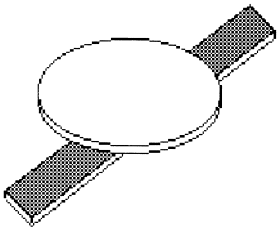


Figure 21.7
Continuation. We would normally assume one bar beneath the disk.

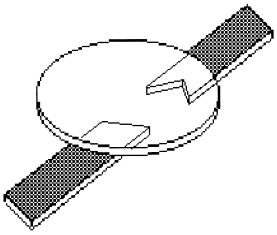


Figure 21.8
Another possible explanation of figure 21.7.

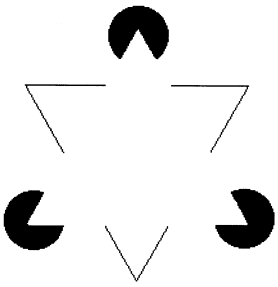


Figure 21.9
More continuation, and some symmetry.

mental context, and if the disk of figure 21.8 is removed to reveal two bars, the infant registers surprise (as measured by breathing and heart rate increases). We will discuss more on early infant studies later in this chapter.

The Italian psychologist Gaetano Kanizsa has come up with a number of interesting examples of perceptual completion or *subjective contours*. Figure 21.9 demonstrates this phenomenon: it is difficult not to see a white triangle located at the center of the figure, although no such triangle actually exists. In the external world, the most probable cause for the improbable alignment of the objects is that a white triangle is lying on top of these objects, covering some and partially masking others.

This phenomenon can also be demonstrated in the auditory domain. Al Bregman has demonstrated this with sinusoidal tones that sweep up and down

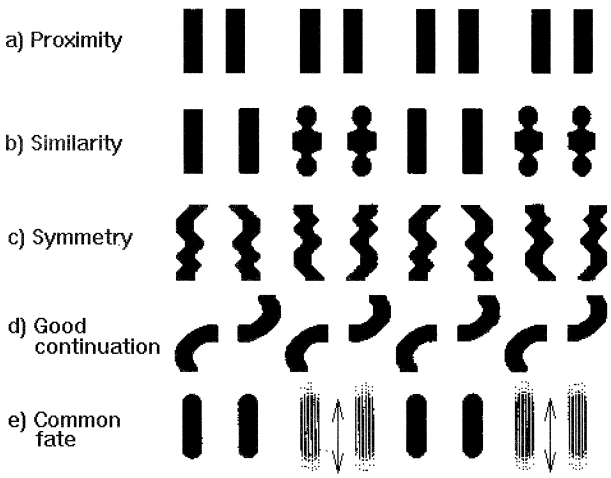


Figure 21.10
Gestalt grouping principles.

in frequency. These tones are interrupted with blank spaces, which cause quite obvious perceptual breaks. When the gaps are masked with bursts of white noise—just as the gaps in the inferred solid bar of figure 21.8 are masked by the disk—the listener makes the inference that the sinusoidal sweeps are continuous. The resulting perception is that a smoothly sweeping sinusoidal sound is occasionally covered up by noise bursts, not that the parts of the sinusoidal sound are actually replaced by bursts of noise, which is what is happening in the signal. The same thing can be done with music: the gaps sound like they are caused by a loose connection in a circuit somewhere; but when the noise bursts fill in the gaps, the illusion is that the music continues throughout.

21.6 The Gestalt Grouping Principles

According to Max Wertheimer, one of the three principal founders of Gestalt psychology, Gestalt principles of grouping are used by the brain when parsing sensory input into objects in the world, especially when information is incomplete or missing altogether. Following are the Gestalt principles of grouping, which are all based on Helmholtz's concept of unconscious inference.

> *Proximity*. Things that are located close together are likely to be grouped as being part of the same object. Figure 21.10a shows the principle of grouping by proximity.
> *Similarity*. When objects are equally spaced, the ones that appear similar tend to be grouped as being related. If objects are similar in shape they are most probably related. (See figure 21.10b.)
> *Symmetry*. Because random unrelated objects in the world are not expected to exhibit symmetry, it would be most improbable for unrelated objects to exhibit symmetric relationships. Figure 21.10c shows principles of both symmetry and similarity.

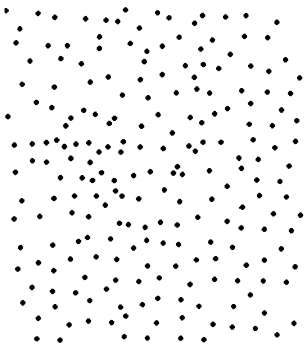


Figure 21.11
Common fate: some "random" dots. Photocopy figure 21.12 onto a transparency sheet, then lay it over figure 21.11. Slide the transparency slightly back and forth, and you will see a woman appear from the field of dots.

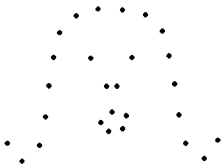


Figure 21.12
Some "random" dots. See figure 21.11.

Good continuation. If objects are collinear, or arranged in such a way that it appears likely that they continue each other, they tend to be grouped perceptually. Figure 21.10d shows the principle of good continuation.

The principles of proximity, similarity, symmetry, and good continuation are considered *weak* principles of grouping, and are often used when the information is incomplete or noisy, or the perceiver has little to go on except the sensory input.

The principle of *common fate* (figure 21.10e) is much stronger. Common fate dictates that objects that move together are likely to be connected. In the world, it is extremely improbable that two things move in a perfectly correlated way unless they are in some way connected. For example, figure 21.11 shows a field of dots, and figure 21.12 shows another field of dots. If figure 21.12 is superimposed over figure 21.11 and moved back and forth, the face shape emerges from the random field of dots, made apparent by the fact that the dots that compose the face move together, and the others do not move.

Demonstrations of auditory common fate typically involve common onset time, common amplitude modulation, and common frequency modulation. One such example involves the grouping of partials and harmonics of a source: we are able to isolate the voice of a speaker or the musical line of a solo instrument in a complex auditory field. The task of isolating a sound source is essentially one of grouping the harmonics or partials that make up the sound; this is done

by grouping those partials by the principle of common fate. The partials tend to move in ensemble, in both frequency and amplitude, and are thus recognized as being part of one object. Individual voices, even though they may be singing the same note, exhibit microfine deviations in pitch and amplitude that allow us to group the voices individually.

Chowning's (1999) examples demonstrate grouping sound sources by common fate. One such demonstration involves a complex bell-like sound consisting of many inharmonic partials. The partials were computer-generated in such a way that they can be grouped into three sets of harmonic partials, each making up a female sung vowel spectrum. When the three voice sets are given a small amount of periodic and random pitch deviation (vibrato), the bell sound is transformed into the sound of three women singing. When the vibrato is removed, the three female voices merge again to form the original bell sound. This is another example of how common fate influences perception of sound.

There are styles of singing in which the vibrato is suppressed as much as possible. Such singing has quite a different effect than typical Western European singing; when the singers are successful in suppressing the vibrato to a sufficient extent, the chorus sound is replaced by a more instrumental timbre. The percept is not one of a group of singers but of a large, complex instrument.

The grouping principles discussed here are actually "wired into" our perceptual machinery. They do not have to be learned by trial and possibly fatal error, because they generally hold in the real world. For example, Elizabeth Spelky did work with early infant development and found that the principle of common fate is used by very young infants. She presented infants with displays of three-dimensional objects and moved some of them together. The infants registered surprise (measured physiologically) when they were shown that the objects that were moving together were not actually parts of the same object, but were artificially caused to move in synchrony. The infants were thus making an unconscious inference based on common fate and good continuation.

We have seen how the Gestalt grouping principle of common fate applies in both vision and audition. Some other Gestalt principles—those of similarity and proximity, for example—might apply to auditory stimuli, and in particular to musical events.

References

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, Mass.: MIT Press.

Chowning, J. C. (1999). "Perceptual Fusion and Auditory Perspective." In Perry R. Cook, ed., *Music, Cognition, and Computerized Sound*, 261–275. Cambridge, Mass.: MIT Press.

Flavell, J. H., and E. M. Markman, eds. (1983). *Cognitive Development*. New York: Wiley.

PART XI

Expertise

Chapter 22

Prospects and Limits of the Empirical Study of Expertise: An Introduction

K. Anders Ericsson and Jacqui Smith

Research on expertise may be one of the most rapidly expanding areas within cognitive psychology and cognitive science. Typically, when a topic becomes popular in psychology, the research approach and the methodology associated with it are also accepted, and the pressure to demonstrate the utility and feasibility of the approach diminishes. Efforts are directed instead toward the theoretical integration of research findings. Furthermore, popularity of a new approach nearly always means that many investigators will adopt it. An even larger number of investigators, however, will adopt only the terminology and will attempt to modify other research approaches to encompass the new concepts. That, in turn, leads to diffusion of the defining characteristics of the "new" approach, making straightforward attempts to integrate published research findings difficult. Because of this process of diffusion, often the new approach will no longer be readily distinguishable from previous alternative research approaches.

In this chapter we attempt to provide a conceptual framework for distinguishing important characteristics of the *original expertise approach*. Our chapter consists of three sections. The first section attempts to characterize the study of expertise in the most general and domain-independent manner so that we can compare the expertise approach with a number of alternative approaches that had similar objectives. The focus of this section is on briefly reviewing some of the outcomes and failures of the earlier approaches. Our goal is to show that the expertise approach can account for these failures at the expense of greater empirical and theoretical complexity. In the second section we specify the nature of the original expertise approach and methodology. Here the pioneering work on chess expertise by de Groot (1978) and Chase and Simon (1973) is used to exemplify the sequence of research steps that characterized the original expertise approach. In the final section we elaborate criteria for these steps and use these criteria to discuss and review the prospects for, and limits of, more recent research on expertise.

Definition of Outstanding Performance and Expertise: A Comparison

On the most general level, the study of expertise seeks to understand and account for what distinguishes outstanding individuals in a domain from less

From chapter 1 in *Toward a General Theory of Expertise*, ed. K. A. Ericsson and J. Smith (New York: Cambridge University Press, 1991), 1–38. Reprinted with permission.

outstanding individuals in that domain, as well as from people in general. We deliberately use the vague term "outstanding" because by not specifying more detailed criteria we are able to point to a number of distinctly different scientific approaches that have addressed the same problem.

In nearly all human endeavors there always appear to be some people who perform at a higher level than others, people who for some reason stand out from the majority. Depending on the historical period and the particular activity involved, such individuals have been labeled exceptional, superior, gifted, talented, specialist, expert, or even lucky. The label used to characterize them reflects an attribution of the major factor responsible for their outstanding behavior, whether it is intended to or not. Scientific efforts to understand the sources of such outstanding behavior have been guided by similar conceptions and attributions.

We limit our discussion to those cases in which the outstanding behavior can be attributed to relatively stable characteristics of the relevant individuals. We believe that stability of the individual characteristics is a necessary condition for any empirical approach seeking to account for the behavior with reference to characteristics of the individual. This constraint does not distinguish whether the characteristics are inherited or acquired. It does, however, eliminate a large number of achievements due to unique immediate environmental circumstances.

The most obvious achievements to be excluded by the stability constraint are those that involve events of fair games of chance, such as winning a large amount of money in a single lottery. More interestingly, the same criterion rules out achievements that occur only once in a lifetime, such as a single scientific discovery, a major artistic creation, a historically significant decision or prediction, or a single victory in a sport. This, of course, does not mean that we reject the possibility of defining criteria for outstanding performances in the arts, sciences, and sports arenas. It does mean, however, that a single achievement in a unique situation does not allow us to infer that the achievement was solely due to the particular individual's characteristics.

In order to support an attribution to the stable characteristics of a person, ideally one would require a series of outstanding achievements under different circumstances. Furthermore, one would like to have a larger group of other individuals (a "control" group of sorts) who have experienced similar opportunities to make contributions or to achieve. In the case in which many other individuals would be equally likely to achieve in similar situations, there is no need to attribute the achievement to special personal characteristics. Almost by definition the numbers of individuals given opportunities in some life realms to achieve and to stand out from the majority are small (e.g., heads of state, army generals, people with vast economic resources). In such cases, even a stable series of achievements cannot unambiguously be linked to stable personal characteristics, because of the confounding influence of a unique stable situation.

Examination of our simple stable-characteristic constraint indicates that many achievements popularly acknowledged as evidence for expertise must be questioned and carefully scrutinized. Another important consequence of this constraint is more indirect and concerns the validity of social evaluation and perception of outstanding performance or ability. One would expect social

evaluation to be greatly influenced by observations of previous performances (not all by the same individual) occurring under unique circumstances. A social judgment, then, might not be the most precise evaluation of an individual's current ability to perform. Ideally, one needs to determine the unique situation of the individual and to observe performances in standardized situations that allow interindividual comparisons (e.g., laboratory tasks or tests). Once it is possible to measure superior performance under standardized conditions, there is no need to rely on social indicators. Attuned to some of the difficulties of definition and assessment, let us now proceed to discuss some scientific approaches that have been directed toward accounting for outstanding or superior performance.

Scientific Approaches to Accounting for Outstanding Performance

Several different scientific approaches have been used to investigate outstanding performance. The constructs that have been investigated have primarily reflected popular attributions regarding the source of the outstanding behavior. These conceptualizations, in turn, have directly influenced what empirical evidence has been considered and collected. Table 22.1 summarizes the different types of stable personal characteristics that have been hypothesized to underlie outstanding performance and links those attributions to associated theoretical constructs and research methods. The attributed personal characteristics noted in table 22.1 reflect a basic belief that behavior either is predominantly influenced by inherited qualities or is a function of learning and acquisition. Further, outstanding performance is attributed either to some general characteristic of the individual or to a specific aspect. The associated theoretical constructs and methodologies reflect these dimensions: *inherited* versus *acquired*, *general* versus *specific*. So, for example, the researcher will focus either on the effects of general traits (e.g., intelligence, personality), specific abilities (e.g., musical ability, spatial ability), and general life and educational experience (e.g., language, study strategies) or on domain-specific training and practice.

One's conception of the likely origins of outstanding performances will greatly influence the group of people selected for study, as well as the type of informa-

Table 22.1
Different approaches to accounting for outstanding performance

Attribution	Construct	Research approach
Primarily inherited		
General abilities	Intelligence, personality	Correlation with personality profile, general intelligence
Specific abilities	E.g., music ability, artistic ability, body build	Correlation with measures of specific ability
Primarily acquired		
General learning and experience	General knowledge and cognitive strategies	Investigation of common processing strategies
Domain-specific training and practice	Domain- or task-specific knowledge	Analysis of task performance, i.e., the expertise approach

tion sought concerning these individuals. For example, investigators pursuing an account in terms of general inherited capacities would be likely to consider individuals regardless of their domains and would be particularly interested in information allowing assessment of the genetic contribution. A longitudinal study of individuals identified as having exceptionally high intelligence, by Terman and his associates (Oden, 1968; Stanley, George, & Solano, 1977; Terman & Oden, 1947), illustrates this approach. A focus on domain-specific acquired characteristics would lead investigators to constrain themselves to one domain or task and to try to assess what was acquired (e.g., specific memory strategies), as well as the process of acquisition.

On a priori grounds one can argue that the most parsimonious theoretical account of outstanding performance is in terms of general, predominantly inherited characteristics. Indeed, in the history of scientific research on superior performance, that approach was initially preferred. It was primarily because of inability to explain certain empirical observations that accounts based on more specific abilities and acquired characteristics came to be seriously considered. We shall briefly consider some of those failures before turning to a consideration of the expertise approach that exemplifies the belief that specific acquired characteristics underlie outstanding performance.

Accounts in Terms of General and Specific Inherited Characteristics

If one wants to attribute outstanding performance to general inherited characteristics, it is reasonable to rely on readily available criteria to identify instances of outstanding behavior and of individuals who exhibit that behavior, criteria such as social evaluation and recognition by one's peers. In the first major study in that area, Galton (1869) used social recognition to identify eminent individuals in a wide range of fields and then studied their familial and genetic origins. Galton argued that individuals gained eminence in the eyes of others because of a long-term history of achievement. Such achievement, he suggested, was the product of a blend of intellectual (natural) ability and personal motivation. He reported strong evidence for eminence's being limited to a relatively small number of families stemming from common ancestors, and he inferred that eminence was genetically determined.

Contemporary work in Galton's time and subsequent studies were directed at uncovering the loci of individual differences in general ability. The genetic nature of those general capacities led investigators to search for differences in basic characteristics of processes, such as the speed of mental processes as reflected by reaction time. In subsequent studies, however, individual differences in performance of simple tasks showed disappointingly low correlations, both among tasks and between performance and indices of ability, such as grade in school (Guilford, 1967).

More recent effort to uncover general basic cognitive processes that could account for individual differences have been inconclusive (Baron, 1978; Carroll, 1978; Cooper & Regan, 1982; Hunt, 1980). For example, research on individual differences in general memory ability has found low correlations of memory performance across different types of material and methods of testing, leading investigators to reject the idea of a general memory ability (Kelley, 1964). More direct evidence against stable basic memory processes comes from repeated

demonstrations that memory performance for specific types of material can be drastically improved even after short periods of practice (Ericsson, 1985; Kliegl, Smith, & Baltes, 1989). Moreover, as Cooper and Regan noted (1982, p. 163), inadequacies in the definition and design of both cognitive tasks and intelligence measures create serious problems for interpreting correlations between measures of basic cognitive processes and ability.

Tests measuring general intelligence have been extremely useful for prediction and diagnosis in a wide range of situations, although there is considerable controversy about what they actually measure (Resnick, 1976; Sternberg, 1982). IQ tests, however, have been remarkably unsuccessftul in accounting for individual differences in levels of performance in the arts and sciences and advanced professions, as measured by social indicators (e.g., money earned, status) and judgments (e.g., prizes, awards) (Tyler, 1965).

There were other lines of research that examined subjects with reliably superior performances and compared them with control groups. Much of that research was similarly motivated by the belief that exceptionally high levels of performance would reflect some basic exceptional ability involving attention (power or concentration), memory, general speed of reaction, or command of logic. Some investigators, however, focused on other stable individual characteristics, such as features of personality, motivation, and perceptual style (e.g., Cattell, 1963; Roe, 1953).

In the 1920s, three Russian professors examined the performance of eight grand masters (world-class chess players) on a wide range of laboratory tests for basic cognitive and perceptual abilities (de Groot, 1946/1978). Surprisingly, the grand masters did not differ from control subjects in those basic abilities, but they were clearly superior in memory tests involving chess positions.

In the case of exceptional chess performance, superior spatial ability often is assumed to be essential (Chase & Simon, 1973; Holding, 1985). Doll and Mayr (1987) compared the performances of about thirty of the best chess players in what was then West Germany with those of almost ninety normal subjects of similar ages, using an IQ test with seven subscales. Only three of the subscales showed reliable differences, and somewhat surprisingly the largest difference between the two groups concerned higher scores for numeric calculation for the chess masters. Doll and Mayr (1987) found no evidence that chess players were selectively better on spatial tasks. In accounting for the unexpected superiority of the chess players on two of the subscales, Doll and Mayr (1987) argued that one reason could be that elite chess players had prior experience in coping with time pressure because of their past chess competitions. When the analysis was restricted to the group of elite chess players, none of the subscales of the IQ test was found to have a reliable correlation with chess-playing performance.

Of the research that has focused not on intelligence but on other relatively stable characteristics of individuals, that by Cattell (1963; Cattell & Drevdahl, 1955) is probably the best example. Cattell sought to determine whether the personality profiles for eminent researchers in physics, biology, and psychology could be distinguished from those of teachers and administrators in the same fields and from those of the general population. Compared with all other groups, top researchers were found to exhibit a consistent profile, being more self-sufficient, dominant, emotionally unstable, introverted, and reflective. Such

a profile supports Galton's earlier opinion that eminence and outstanding achievement in a field are products not only of ability but also of aspects of personal motivation. Motivation and striving for excellence often are focused on a small number of domains or even a single domain, suggesting that aspects of motivation may well be acquired.

Despite these hints at possible personality patterns, the research approach of accounting for outstanding and superior performance in terms of general inherited characteristics has been largely unsuccessful in identifying strong and replicable relations. The search for links to specific inherited abilities has been similarly inconclusive. Indeed, as the specific characteristics proposed to account for the superior performance become integral to that performance, it becomes difficult to rule out the possibility that such characteristics have not been acquired as a result of many years of extensive training and practice. Investigators have therefore focused their attention on characteristics that appear in children and that reflect basic capacities for which a genetic origin is plausible. We shall briefly consider two examples of such basic capabilities, namely, absolute pitch among musicians and physiological differences among elite athletes.

A recent review of the research on absolute pitch shows that most of the empirical evidence favors an account in terms of acquired skill (Ericsson & Faivre, 1988). The ability to recognize musical pitch is not an all-or-none skill, and many musicians have it to various degrees. They display the best performance on their own instruments, and their performance decreases as artificial tones from a tone generator are presented (Bachem, 1937). The ability to name pitches correctly is closely related to the amount of one's formal musical training (Oakes, 1955). Furthermore, pitch recognition can be dramatically improved with training, and one musician has documented how he acquired absolute pitch through long-term training (Brady, 1970).

Similarly, a recent review shows that many anatomical characteristics of elite athletes, such as larger hearts, more capillaries for muscles, and the proportions of different types of muscle fibers, are acquired during years of practice (Ericsson, 1990). Such findings showing the far-reaching effects of training do not, however, rule out possible genetic constraints. An individual's height and overall physique are determined by genetic factors (Wilson, 1986). Height and physique, for example, impose important constraints in many physical and sports domains, such as basketball, high jumping, gymnastics, ballet, and professional riding. It is also conceivable that genetic factors might influence the rate of improvement due to training. Nevertheless, training and preparation appear to be necessary prerequisites and important determinants of outstanding performance. We turn to a brief discussion of accounts of outstanding and superior performance based on acquired characteristics.

Accounts in Terms of Specific Acquired Characteristics: The Expertise Approach

In this brief review we have seen that the more parsimonious theoretical approaches relying on stable inherited characteristics seem inadequate to account for outstanding and superior performance. It is therefore necessary to consider accounts based on acquired characteristics. Here we need to identify not only what the acquired characteristics are but also the process by which they are acquired.

How long is the acquisition period, and over what time frame do we need to observe and monitor changes in performance? Simon and Chase (1973) were the first to observe that 10 years or more of full-time preparation are required to attain an international level of performance in chess. Studies by Hayes (1981) and Bloom (1985) revealed that a decade of intensive preparation is necessary to become an international performer in sports or in the arts or sciences. In a recent review, Ericsson and Crutcher (1990) found consistent support for the requirement of 10 years of intensive preparation in a wide range of studies of international levels of performance. Furthermore, Ericsson and Crutcher (1990) found for many domains that most international-level performers had been seriously involved in their domains before the age of 6 years. The period of preparation for superior performance appears to cover a major proportion of these individuals' development during adolescence and early adulthood.

A detailed analysis of acquisition processes extending over decades under widely different environmental circumstances is extraordinarily difficult to conduct. Without a theoretical framework to outline the relevant aspects, the number of possible factors that could be critical to attain superior performance is vast. One can, of course, gain some idea of the range of factors by reading biographies and analyses of unusual events or circumstances in the lives of outstanding scientists and artists (Albert, 1983; McCurdy, 1983). It is unlikely, though, that descriptive studies seeking correlations between ultimate performance of individuals and information about their developmental histories will ever be able to yield conclusive results. A much more promising approach is offered by a careful analysis of the attained performance. This is the crux of the expertise approach.

The expertise approach differs from the approaches discussed earlier in some important respects. The other approaches were attempts to measure independently the constructs hypothesized to be the sources and bases of outstanding performance. In contrast, the expertise approach is an attempt to describe the critical performance under standardized conditions, to analyze it, and to identify the components of the performance that make it superior.

Two features distinguish the expertise approach from other approaches: first, the insistence that it is necessary to identify or design a collection of representative tasks to capture the relevant aspects of superior performance in a domain and to elicit superior performance under laboratory conditions; second, the proposal that systematic empirical analysis of the processes leading to the superior performance will allow assessment of critical mediating mechanisms. Moreover, it is possible to analyze the types of learning or adaptation processes by which these mechanisms can be acquired and to study their acquisition in real life or under laboratory conditions.

The expertise approach is more limited in its application than the other approaches reviewed earlier. Whereas the other approaches can use social indicators as criterion variables of outstanding performance, the expertise approach requires the design of a set of standardized tasks wherein the superior performance can be demonstrated and reliability reproduced. With this important limitation in mind, we now turn to a closer examination of the original expertise approach.

The Original Expertise Approach: The Pioneering Work on Chess

There is no consensus on how the expertise approach should be characterized. If one takes the original work on chess expertise by de Groot (1978) and Chase and Simon (1973), however, it is possible to extract three general characteristics. First, the focus is on producing and observing outstanding performance in the laboratory under relatively standardized conditions. Second, there is a theoretical concern to analyze and describe the cognitive processes critical to the production of an outstanding performance on such tasks. Finally, the critical cognitive processes are examined, and explicit learning mechanisms are proposed to account for their acquisition.

If one is interested in reproducing superior performance under standardized conditions, one should give preference to domains in which there are accepted measures of performance. Chess provides such a domain. It is possible to measure an individual's chess-playing ability from the results of matches against different opponents in different tournaments (Elo, 1978). It is easy to select groups of chess players who differ sufficiently in chess ability that the probability of one of the weaker players beating one of the stronger players in a particular game is remote.

A critical issue in the expertise approach is *how to identify standardized tasks* that will allow the real-life outstanding performance to be reproduced in the laboratory. Because of the interactive nature of chess games and the vast number of possible sequences of moves, the same sequences of chess moves are hardly ever observed in two different chess games. Better chess players will consistently win over weaker chess players employing a wide variety of chess-playing styles. One could therefore argue that the better chess players consistently select moves as good as, or better than, the moves selected by weaker players. De Groot (1978) argued that it is possible to develop a collection of well-defined tasks capturing chess expertise by having chess players select the "best next move" for a number of different chess positions. Measurement of performance in this task requires that it be possible to evaluate qualitatively, on a priori grounds, the dependent variable, that is, the next chess move selected for a given chess position. It is not currently possible to evaluate the quality of chess moves for an arbitrary chess position. In fact, one international chess master claims to have spent a great part of his life unsuccessfully seeking to determine the best move for one particular chess position (Saariluoma, 1984).

De Groot (1978) collected think-aloud protocols from chess players of widely differing levels of expertise while they selected their best next moves for several chess positions. After extended analysis of these classic positions, however, he found that only *one* of them differentiated between grand masters and other chess experts who differed greatly in chess ability: All of the very best chess players selected better moves than did any of the comparatively weak players (nonoverlapping). Hence, he inferred that the task of selecting moves for that chess position must elicit cognitive processes that differentiate chess players at different levels of expertise.

Another pioneering aspect of de Groot's study was his use of verbal protocols. He was able to localize differences in cognitive processes between the grand masters and the other class experts by analyzing think-aloud protocols from his best-next-move task. He found that both masters and experts spent

about 10 minutes before deciding on a move. In the beginning, the players familiarized themselves with the chess position, evaluated the position for strengths and weaknesses, and identified a range of promising moves. Later they explored in greater depth the consequences of a few of those moves. On average, both masters and experts considered more than thirty move possibilities involving both Black and White and considered three or four distinctly different first moves.

De Groot (1978) first examined the possibility that, compared with chess experts, the grand masters were able to explore longer move combinations and thereby uncover the best move. He found, however, that the maximum depth of the search (i.e., the length of move combinations) was virtually the same for the two groups. When de Groot then focused his analysis on how the players came to consider different moves for the position, he did find differences. Few of the chess experts initially mentioned the best move, whereas most of the grand masters had noticed the best move during the familiarization with the position. More generally, de Groot argued, on the basis of his analysis of the protocols, that the grand masters perceived and recognized the characteristics of a chess position and evaluated possible moves by relying on their extensive experience rather than by uncovering those characteristics by calculation and evaluation of move possibilities. In some cases the discovery of promising chess moves was linked to the verbal report of a localized weakness in the opponent's chess position. Other grand masters discovered the same move without any verbal report of a mediating step (de Groot, 1978, p. 298). The superior chess-playing ability of more experienced chess players, according to de Groot, is attributable to their extensive experience, allowing retrieval of direct associations in memory between characteristics of chess positions and appropriate methods and moves. De Groot (1978, p. 316) argued that mastery in "the field of shoemaking, painting, building, [or] confectionary" is due to a similar accumulation of experiential linkings.

To examine the critical perceptual processing occurring at the initial presentation of a chess position, de Groot (1978) briefly showed subjects a middle-game chess position (2–10 seconds). Shortly after the end of the presentation the chess players gave retrospective reports on their thoughts and perceptions during the brief presentation and also recalled the presented chess position as best they could. From the verbal reports, de Groot found that the position was perceived in large complexes (e.g., a pawn structure, a castled position) and that unusual characteristics of the position (such as an exposed piece or a far-advanced pawn) were noticed. Within this brief time, the chess masters were found to integrate all the characteristics of the position into a single whole, whereas the less experienced players were not able to do so. The chess masters also often perceived the best move within that short exposure time. The analysis of the amount recalled from the various chess positions was consistent with the evidence derived from the verbal reports. Chess masters were able to recall the positions of all the 20–30 chess pieces virtually perfectly, whereas the positions recalled by the less experienced chess experts ranged from 50 to 70 percent.

The classic study of Chase and Simon (1973) followed up on this superior memory performance by chess masters for briefly presented chess positions. They designed a standardized memory task in which subjects were presented

with a chess position for 5 seconds with the sole task of the subjects being to recall the locations of as many chess pieces as possible. We shall later review more carefully to what extent this new task can be viewed as capturing the cognitive processes underlying superior chess-playing performance.

With that memory task, Chase and Simon (1973) were able to corroborate de Groot's earlier finding that chess players with higher levels of expertise recalled the correct locations of many more pieces for representative chess positions. They also went a significant step farther and experimentally varied the characteristics of the presented configurations of the chess pieces. For chessboards with randomly placed pieces, the memory performances of the chess masters were no better than those of novice chess players, showing that the superior memory performance of the master depends on the presence of meaningful relations between the chess pieces, the kinds of relations seen in actual chess games.

Chase and Simon (1973) found that a player's ability to reproduce from memory the previously presented chess position proceeded in bursts in which chess pieces were rapidly placed, with pauses of a couple of seconds between bursts. The pieces belonging to a burst were shown to reflect meaningfully related configurations of pieces (i.e., chunks) that corresponded well to the complexes discovered by de Groot (1978). The chess masters were found to differ from other chess players primarily in the number of pieces belonging to a chunk, that is, the size of the chunk. In support of the hypothesis that memory and perception of chess positions rely on the same encoding processes, Chase and Simon (1973) demonstrated that the recall process had a structure similar to that of the process of reproducing perceptually available chess positions. Rather than discuss the large number of additional empirical studies by Chase and Simon (1973), we shall change the focus and consider their theoretical effort to specify the detailed processes underlying superior memory performance and the relation of these processes to general constraints on human information processing.

One of the most severe constraints on an account that is based on acquired knowledge and skill involves explicating what has been acquired and showing that the acquired characteristics are sufficient to account for the superior performance without violating the limitations of the general capacities of human information processing (Newell & Simon, 1972). The superior recall of 15–30 chess pieces by chess masters would at first glance seem to be inconsistent with the limited capacity of short-term memory in humans, which allows storage of around 7 chunks (Miller, 1956). Chase and Simon (1973) found that the number of chunks recalled by chess players at all skill levels was well within the limit of around 7 ± 2. They attributed the difference in memory performance between strong and weak players to the fact that the more expert chess players were able to recognize more complex chunks, that is, chunks with a larger number of chess pieces per chunk.

On the basis of computer simulations of the encoding and recall of middlegame chess positions, Simon and Gilmartin (1973) were able to show that 1,000 chunks were sufficient to reproduce the memory performance of a chess expert. They estimated that simulation of the performance of a chess master would require between 10,000 and 100,000 chunks. Assuming that the superior perfor-

mance of the expert depends on the recognition of familiar patterns that index previously stored relevant knowledge of successful methods (actions), the time-consuming process of becoming an expert would consist in acquiring those patterns and the associated knowledge. Simon and Chase (1973) estimated that around 3,000 hours are required to become an expert, and around 30,000 hours to become a chess master. They also commented that "the organization of the Master's elaborate repertoire of information takes thousands of hours to build up, and the same is true of any skilled task (e.g., football, music). That is why *practice* is the major independent variable in the acquisition of skill" (p. 279). Whether or not one agrees with the Chase and Simon theory of expertise, it would be unwise to confound the methodology of their research with the theoretical assumptions of their specific theory. Indeed, Chase and Simon (1973) were rather cautious when they proposed their theory, describing it as simply a rough first approximation.

The Three Steps of the Original Expertise Approach

From our review of the pioneering research on chess expertise we have extracted three steps. The first step involves capturing the essence of superior performance under standardized laboratory conditions by identifying representative tasks. In the following sections we try to distinguish between collections of tasks that capture the superior performance and collections of tasks that measure a related function or ability. In our review of the initial work on chess, we argued that only the task that required that subjects consistently select the "best moves" meets the criterion of capturing the nature of superior performance. Two other tasks, one involving perception and the other measuring memory for briefly presented chess positions, assess related functions but do not directly represent chess-playing skill.

The second step involves a detailed analysis of the superior performance. The pioneering research on chess nicely illustrates the use of refined analyses of sequences of verbal reports and placement of chess pieces to infer the underlying cognitive processes mediating the superior performance, as well as the use of experimental manipulation of stimulus materials.

The third and final step involves efforts to account for the acquisition of the characteristics and cognitive structures and processes that have been found to mediate the superior performances of experts. A persistent failure to identify conditions under which the critical characteristics could be acquired or improved would provide strong evidence that those characteristics are unmodifiable and hence basic and most likely inherited.

Our explication of the original expertise approach imposes clear limits for its successful application. Unless the essence of the superior performance of the expert can be captured in the laboratory (satisfying the criterion for the first step), there will not be a performance to be further analyzed in terms of its mediating processes. Similarly, failure to identify mediating processes that can account for the superior performance during the second step will leave the investigator with only the original differences in overall performance and will make the third step essentially superfluous.

At the same time, our explication of the expertise approach is applicable to any phenomenon involving reliably superior performance that can be captured in the laboratory. We believe that an attempt to encompass phenomena normally labeled as perceptual (e.g., chicken sexing), motoric (e.g., typing), or knowledge-based (e.g., physics) within the same overall approach will allow us to identify common methodological and theoretical issues and to consider a common and more differentiated set of learning mechanisms in accounting for achievement of superior performance in any one of these different domains. Such an approach will have the additional advantage of allowing us to consider the many different perceptual, memory, motoric, and knowledge-based aspects of superior performance in domains like chess (Charness, 1991), physics (Anzai, 1991), medicine (Patel & Groen, 1991), performing arts and sports (Allard & Starkes, 1991), and music (Sloboda, 1991).

Capturing Superior Performance: The First Step

The first step in the expertise approach involves finding or designing a collection of tasks to capture the superior performance in the appropriate domain. If one is able to identify such a collection of tasks, the following important advantages will accrue: First, the performance of the designed tasks will reflect the stable characteristics of the superior real-life performance. More important, the availability of such a collection of tasks will allow us to study the performance of the experts extensively in order to accumulate sufficient information on the mediating processes to make a detailed assessment and analysis. During these extensive observations of performance, we should not expect significant changes due to learning and practice, as we shall be monitoring stable processes that have been adapted and perfected over a long period of time. The period during which performance will be observed will be negligible in comparison.

Finally, these collections of tasks will provide us with an excellent testing ground for studying how rapidly the various identified characteristics can be acquired through practice. In fact, one could argue that with an adequate collection of tasks, the rates of acquisition should be comparable for practice with the collection of tasks and in real life. If, on the other hand, the collection of designed tasks does not elicit the mechanisms that mediate superior real-life performance, or does so only partially, then we are likely to see substantial learning and changes in the processes as a result of further practice. Collections of tasks that lead to rapid rises in levels of performance by experts with further practice are unlikely to yield an adequate representation of superior performance. Even more devastating evidence against the claim that such a collection can capture superior performance comes from situations in which novices have matched or surpassed the performance levels of experts after only a few weeks or months of practice.

For some types of expertise it is easy to identify such a collection of tasks, but in most cases it is the most difficult step. We shall first describe some simple cases and then turn to the difficult issues involved in designing a collection of tasks to characterize real-life expertise. We shall also consider the advantages and problems of designing a collection of memory tasks to study superior memory performance by experts, as opposed to studying directly the superior performance of experts.

Tasks Capturing Real-Life Expertise. There are few instances of real-life expertise in which superior performance can be demonstrated under relatively standardized conditions. Mental calculators and memory experts provide such instances. They often exhibit their performance under conditions similar to those used in traditional experiments. In both of these cases it is easy to define a large pool of different stimuli (e.g., 10 billion possible multiplications of two 5-digit numbers, or 100 trillion digit sequences of 14 digits). Drawing on this pool of items, the experimenter can observe the performance in a large number of different trials and accumulate information on the cognitive processes underlying the expertise. Similarly, some types of psychomotor performance, such as typing, and some sporting events can easily be imported into the laboratory.

Apart from the preceding cases, the design of standardized tasks to capture real-life expert performance is difficult. The problem is somewhat similar to that of isolating phenomena in the natural and biological sciences. By careful analysis of the expert performance in real life, we try to identify recurrent activities that can be reproduced under controlled conditions. In those domains in which expertise can be measured, it is important to restrict the focus to those activities that are involved in producing the relevant performance or resulting product. One should search for goal-directed activities that result in overt behavior that can be reproduced by presentation of the appropriate stimuli.

A nice illustration of this procedure comes from the previously described research on chess, in which de Groot (1978) designed the task of selecting the best next move for a given middle-game position. It should be possible to collect a large number of such positions with which even top-level chess players would be unfamiliar. In extracting out a single chess position from a chess game, one is faced with a problem that is common in research on expertise, namely, the determination of the correct response, or the reliable evaluation of selected moves. Given that currently there was no method available that could have provided that information objectively, de Groot (1978) spent an extended period carefully analyzing the selected chess position to evaluate the relative merits of different moves. A different method of dealing with this problem was offered in a recent study by Saariluoma (1984), who selected chess positions that had clearly discernible best next moves. Both of these methods are oriented toward finding or designing a small set of tasks, and they cannot easily be extended into specifying a large population of tasks that could be claimed to capture the chess expertise.

In most other complex task domains, such as physics and medical diagnosis, investigators tend to select a small number of tasks without specifying the population from which those tasks were chosen to be a representative sample. One reason for this is that a detailed task analysis of even a single complex problem is difficult and extraordinarily time-consuming. More important, our knowledge of complex domains of expertise is incomplete, and it would not at this time be possible to specify a population of tasks to capture such expertise. Many scientists, however, are working on building expert systems in which the tasks and prerequisite knowledge must be specified, and other researchers are working on describing the formal characteristics of various task environments. (see Charness, 1991).

In many domains, experts produce complex products such as texts on a given topic or performances of a given piece of music. Although judges can reliably assess the superior quality of the product, it is difficult to analyze such products in order to identify the measurable aspects capturing the superior quality of the product. Hence, in their analysis of expertise in writing, Scardamalia and Bereiter (1991) focus on systematic characteristics of the cognitive processes involved in designing and writing a text in an effort to differentiate expert from novice writers.

It is, of course, possible to give up the hope of designing a collection of tasks that could capture the full extent of the superior performance and focus instead on one or more well-defined activities involved in the expertise or measuring knowledge about the task domain. In adopting such an approach, one no longer can be certain that one is examining cognitive structures and processes essential to the superior performance. Occasionally, expected differences between the performance of novices and that of experts in component activities are not found. For example, Lewis (1981) found no reliable differences in performance on algebra problems between expert mathematicians and the top third of a group of college students. The most frequently studied activity related to expert performance is memory for meaningful stimuli from the task domain.

Tasks Focusing on Domain-Specific Memory Performance. In the context of the difficulties of identifying a collection of tasks that can capture the expertise, it is easy to see the attractiveness of studying memory performance. It is possible to evaluate memory performance for presented information by means of recognition and reproduction of literal details (e.g., correct placement of chess pieces), which does not involve any in-depth analysis of tasks or prior knowledge in the given domain. Large samples of different meaningful stimuli can relatively easily be extracted from a given domain even though no formal description of the corresponding population of stimuli is given. Similarly, it is relatively easy to assemble unrepresentative or even meaningless stimuli by recombining stimulus elements in an arbitrary or random manner.

In a wide range of different domains, experts have been shown to display superior memory performance for representative stimuli from their domains of expertise when adaptations of Chase and Simon's (1973) original procedure have been used: chess (for a review, see Charness, 1991); bridge (Charness, 1979; Engle & Bukstel, 1978); go (Reitman, 1976); music notation (Sloboda, 1976); electronic circuit diagrams (Egan & Schwartz, 1979); computer programming (McKeithen, Reitman, Rueter, & Hirtle, 1981); dance, basketball, and field hockey (Allard & Starkes, 1991). Other studies have shown superior retention of domain-related information as a function of the subject's amount of knowledge of the domain, such as baseball (Chiesi, Spilich, & Voss, 1979; Spilich, Vesonder, Chiesi, & Voss, 1979; Voss, Vesonder, & Spilich, 1980) or soccer (Morris, Gruneberg, Sykes, & Merrick, 1981; Morris, Tweedy, & Gruneberg, 1985). Hence, many studies have found evidence supporting a monotonic relation between recall performance for a domain and expertise in that domain. There are, however, several lines of research that have questioned the generality of that relation. Sloboda (1991) points out the striking similarity in accuracy

and structure of recall of presented melodies between musicians and nonmusicians, which he attributes to shared extensive experience with music. Allard and Starkes (1991) show that superior recall of briefly presented game situations by elite players, as compared with intramural players, is not always found in sports with speed stress, such as volleyball. Finally, Patel and Groen (1991) demonstrate that levels of medical expertise have nonmonotonic relations to the amounts of information recalled from presented medical cases, which they attribute in part to the ability of experts to efficiently identify the information relevant to the medical diagnosis. These findings show that superior memory performance is not an inevitable consequence of attaining expertise.

It is thus questionable that a collection of tasks to measure the superior memory of experts can be claimed to really capture the expertise in question. With the exception of experts on memory tasks, superior performance by experts in many domains does not include explicit tests of memory performance. Moreover, there is no reason to believe that experts explicitly train with the goal of increasing their memory performance. It is therefore unlikely that their memory performance would have reached a stable maximum. We shall later discuss in more detail the cognitive processes relating memory performance and expertise.

An issue shared by studies of superior memory performance and studies of superior performance in other realms is the problem of determining the stimulus characteristics necessary to evoke performance in the laboratory analogous to real-life expertise.

Finding the Appropriate Stimuli to Evoke Superior Performance. In capturing expert-level performance, one attempts to create a situation that is maximally simple and yet sufficiently similar to the real-life situation to allow the reproduction of the expertise under laboratory conditions. The mere demonstration that an expert-level performance can be reproduced under controlled laboratory conditions reveals something important about the mechanisms underlying the corresponding expertise. It reduces the number of possible stimulus variables that are critical to performance, and it can also eliminate a number of systematic covariations that would make the real-life performance much easier than it would initially appear. Despite the critical importance of the process of finding appropriate stimuli to evoke superior performance, that process has rarely been documented. Ericsson and Polson (1988a, 1988b) investigated the ability of expert waiters and waitresses to match meal orders to customers. They reproduced under laboratory conditions the superior memory performance related to dinner orders by simulating actual customers with photos of faces. Similarly, Bennett (1983) reproduced the superior memory performance related to drink orders by cocktail waitresses in a simulated situation with dolls representing customers. Hence, highly schematic stimuli are sufficient to elicit the perceptual and representational mechanisms that mediate superior memory performance. Similarly, Chase and Simon (1973) found that the memory performance of two chess experts did not differ for chessboards with real pieces and schematic diagrams of chess positions, whereas a beginner at chess showed poorer recall with schematic diagrams because of lack of familiarity with the diagram notation. When they exposed a chess expert to an unfamiliar type of

letter diagram representing the chess positions, his memory performance was only half as good as his performance with a real board. But after only 16 trials, his performance with the unfamiliar diagrams had improved to the level of his performance with the real board. Charness (1991) provides a review of the current research using different visual representations of chess positions.

There is some evidence that there are limits to the extent to which stimuli can be abstracted. Gilhooly, Wood, Kinnear, and Green (1988) demonstrated that the lack of superior memory performance by expert map users, as compared with the novices studied by Thorndyke and Stasz (1980), could be attributed to their use of schematic maps (mainly used by tourists) as stimuli. By studying recall of both schematic maps and more advanced contour maps by expert and novice map users, Gilhooly et al. (1988) found, as expected, superior memory of contour maps by the experts, but no differences between experts and novices for the commonly available schematic maps. The fact that superior performances can be reproduced in the laboratory with schematic stimuli is important not only for practical purposes but also for theoretical analyses of the mediating mechanisms.

The issues of how to design representative laboratory tasks are discussed in many chapters in this volume. For example, Patel and Groen (1991) consider the differences between medical diagnoses based on written texts presenting medical cases and diagnoses based on interviews with real patients. Dörner and Schölkopf (chapter 9, this volume) report on the management of simulations of very complex systems.

Summary. The essential first step of the study of expert performance involves identifying a collection of standardized tasks that can capture the superior performance under controlled conditions. It is a necessary condition for further analysis that superior performance by experts be reliably shown for the designed tasks. In complex domains it is often especially difficult to identify a population of tasks to capture the expertise; it may be possible to identify instead a small number of representative tasks to elicit superior performance. Nonetheless, it may be useful to think of expertise in terms of a corresponding population of tasks. Various experts may, however, require different populations of tasks. Patel and Groen (chapter 4, this volume) show that with increasing expertise in medicine, experts become more specialized in particular areas of medicine. Similar specialization is to be expected in most complex domains. To capture specialized expertise adequately, it is necessary to design special populations of tasks appropriate for a small group of experts or even individual experts (case studies). Superior memory performance by an expert is a legitimate subject for study as long as we keep in mind that the processes underlying the superior memory performance may only partially overlap with those that generally underlie the superior performance of experts.

The fact that it is possible to reproduce expert performance in a laboratory task has important theoretical implications. It reduces the significance of large numbers of factors that influence complex real-life situations. Furthermore, it indicates a fair degree of generalizability, especially concerning the detailed stimulus representation. Let us now turn to further analysis of the processes that mediate superior performance.

Analysis of Expert Performance: The Second Step
After identifying collection of tasks that can capture the superior performance of experts, one can apply the full range of methods of analysis in cognitive psychology to examine the phenomena associated with a particular type of expertise. In the following sections we present a brief outline of the wide range of observations that can be made to infer information about the processes mediating superior performance. We then discuss different research paradigms, such as comparisons of performance by experts and novices in a small number of tasks, and extended analysis of individual experts. Finally, we report on analyses of particular types of superior performance, such as superior memory performance.

Performance Analysis: Methods of Inferring Mediating Processes. It is clear that one cannot directly observe mediating cognitive processes, but what can be observed concurrently with cognitive processes can be related to the underlying cognitive processes within the information-processing theory of cognition. Figure 22.1 shows a number of different types of observations that can be collected on any cognitive process. At the top of figure 22.1, cognitive processing is represented schematically as a series of internal processing steps, as proposed by the information-processing theory of human cognition. These internal processing steps cannot, of course, be observed directly, but it is possible to

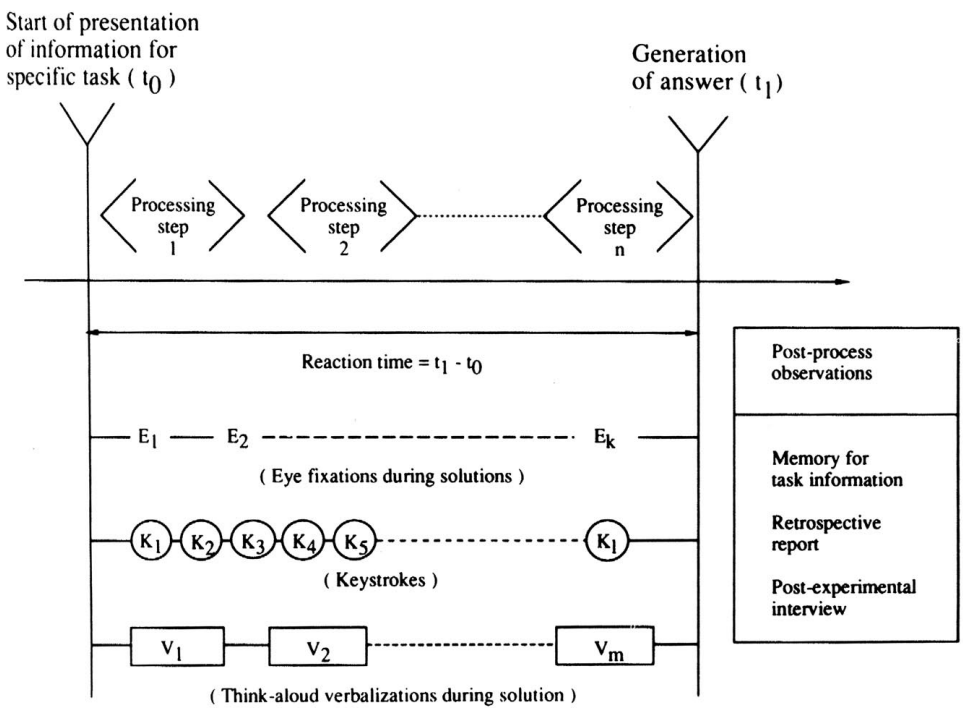


Figure 22.1
An overview of different types of observations on cognitive processes mediating performance on a task, adapted from a figure in Ericsson and Oliver (1988).

specify hypotheses about the relations between the internal processing steps and observable behavior. For example, when a subject fixes his or her gaze on a specific item in a visually presented table of information, we can infer that the corresponding internal steps involve processing that information. On the basis of veridical recall of the presented information after the task has been completed and the presented information is no longer available to the subject, one can infer that that information was processed during the completion of the task. In research on transcription typing, it is possible to determine what part of the text the typist is looking at and what part of the text is simultaneously being typed. The general finding is that the higher the skill level of the typist, the farther ahead in the text the typist looks during typing. Being able to look ahead in the text appears to be critical to the superior typing speeds of expert typists, because when their freedom to look ahead is experimentally restricted, their typing speeds are reduced to levels approaching those for novice typists (Salthouse, 1984, chapter 11, this volume).

It is possible to extend our analysis beyond the processing of presented information and consider one's access of preexisting knowledge and procedures. In that case, a *task analysis* of the particular task should be performed before the data collection. A task analysis involves specifying a number of different sequences of processing steps that could generate the correct answer for a specific task given the subjects' preexisting knowledge. In well-defined task domains, such as mental multiplication or problem solving in logic, it is relatively easy to specify nearly exhaustively the different sequences of processing steps leading to a correct answer in an efficient manner. In more complex domains, the a priori task analysis makes explicit the pool of hypothesized processing sequences that is being considered. On the basis of the think-aloud verbalizations of subjects, one can determine only that the verbalized information was accessed. A task analysis is critical for relating the verbalized information to the underlying cognitive processes leading to its access or generation (Ericsson & Simon, 1984).

Analysis of think-aloud verbalizations is time-consuming, and therefore researchers in expertise using these types of data tend to collect data on many subjects for a small number of tasks (expert–novice comparisons) or to collect data on individual subjects for a large number of tasks (case studies).

Expert–Novice Comparisons. Comparison of think-aloud verbalizations by experts and novices is the best-known method of assessing differences in the mediating processes as functions of the subjects' levels of expertise: Subjects at different levels of expertise are asked to think aloud while carrying out a small number of representative tasks. The number of tasks usually is not sufficient for assessing stable characteristics of individual subjects; the focus is on comparing the groups of subjects to identify salient differences in regard to mediating knowledge and processes.

The types of differences found in a wide range of domains of expertise are remarkably consistent with those originally noted by de Groot (1978) in the domain of chess. Expert performers tend to retrieve a solution method (e.g., next moves for a chess position) as part of the immediate comprehension of the task, whereas less experienced subjects have to construct a representation of the

task deliberately and generate a step-by-step solution, as shown by research on physics problems (Anzai, 1991; Chi, Glaser, & Rees, 1982; Larkin, McDermott, Simon, & Simon, 1980; Simon & Simon, 1978) and algebra-word problems (Hinsley, Hayes, & Simon, 1977). Medical experts generate their diagnoses by studying the symptoms (forward reasoning), whereas less experienced medical students tend to check the correctness of a diagnosis by inspecting relevant symptoms (backward reasoning) (Patel & Groen, 1991).

On the same theme, expert performers have a body of knowledge that not only is more extensive than that for nonexperts but also is more accessible (Feltovich, Johnson, Moller, & Swanson, 1984; Johnson et al., 1981; Voss, Greene, Post, & Penner, 1983). Whenever knowledge is relevant, experts appear to access it efficiently (Jeffries, Turner, Polson, & Atwood, 1981). The experts are therefore able to notice inconsistencies rapidly, and thus inconsistent hypotheses are rejected rapidly in favor of the correct diagnosis (Feltovich et al., 1984; Johnson et al., 1981). On presentation, information in the problem is integrated with the relevant domain knowledge (Patel & Groen, 1986, 1991).

Similar characteristics of expert performance are found across different domains of expertise. The studies cited earlier suggest several important characteristics that can be more effectively studied in relation to tasks particularly designed to elicit them in a more controlled manner. We shall consider such research shortly.

Extensive Case Studies of Single Subjects. In contrast to the group studies discussed earlier, in which small numbers of tasks were used to elicit the cognitive processes of experts, we shall briefly consider two examples of research efforts that have used detailed case studies in order to describe the cognitive processes underlying superior performance.

The first example draws on several case studies of calendar calculations. Calendar calculation is the rather astounding ability to name the day of the week on which a given date falls. For example, when asked on what day of the week August 5, 1934, fell, such a subject would be able to say, correctly, that it was a Sunday. A major interest in this curious ability derives from the fact that several individuals with this skill have been severely mentally retarded, and little is known about how the ability emerged or was acquired. Analysis of this performance is further complicated by the low intelligence of the subjects. On the basis of a task analysis, where no knowledge about calculation can be assumed for these mentally retarded subjects, one is led to assume that the subjects must have memorized the information for all dates.

Investigators have examined a fairly large number of individuals for whom the ability of calendar calculation has been substantiated (for reviews, see Ericsson & Faivre, 1988, and Howe & Smith, 1988). Most calendar calculators can demonstrate that ability for only a limited range of years. All such subjects examined have been unable to explain how they know the correct answers. Some investigators, however, have been able to assess mediating steps by analyzing these subjects' mumblings prior to reporting an answer. Other investigators have been able to obtain informative retrospective reports on mediating steps. The most reasonable conclusion seems to be that the detailed structures of these subjects' processes differ from subject to subject and rely on a

combination of memory for specific dates and some limited specialized calculation (Howe & Smith, 1988). The rare calendar calculators whose abilities extend from A.D. 0 to A.D. 999,999 appear to use a version of known algorithms that can be mastered by a graduate student within a couple of weeks to reach a comparable level of performance (Addis & Parson, described in Ericsson & Faivre, 1988).

A second example of single-subject research to analyze expert performance draws on the many case studies of memory experts. Studies of expert memory performances are particularly suited for the laboratory and can capitalize on the long tradition of experimental research on memory. The same research tradition has primarily used stimuli that have been selected to be meaningless, or at least has minimized the role of knowledge in order to capture basic memory processes. It has, however, been difficult to account for vastly superior memory performance within this tradition, and occasionally investigators have suggested that such exceptional individuals are endowed with structurally different memory systems (Luria, 1968; Wechsler, 1952). Analysis of expert memory performance is difficult even in the information-processing tradition, because it is virtually impossible to conduct an a priori task analysis specifying the mediating processing steps and the relevant knowledge used to store information efficiently in memory.

One of the methods available is to use think-aloud and retrospective verbal reports to identify the knowledge used by an individual memory expert and experimentally evaluate hypotheses about the mediating role of that knowledge. For each individual expert it is possible to hypothesize which stimuli could and could not be successfully encoded using the uncovered mediating knowledge. By comparing memory performances for compatible and incompatible stimuli, it is possible to validate hypotheses about the mediating knowledge using the general method developed by Chase and Simon (1979). In a study of a long-distance runner who acquired an exceptional digit span through extended training, Chase and Ericsson (1981) found that the runner encoded sequences of three digits (513) as familiar running times (5 minutes and 13 seconds in a mile race) whenever possible. When the runner was presented with experimentally prepared sequences of triplets of digits that could not be encoded as running times (483 would be 4 minutes and 83 seconds), his digit-span performance was dramatically reduced, and for prepared sequences of triplets all of which could be encoded as running times, his performance was reliably improved over his performance with random digit sequences. Similarly, Sloboda (1991) shows that superior memory performance for classical music by idiots savants is mediated by knowledge of that type of music and cannot be generalized to modern atonal music.

Case studies of memory experts have revealed that the knowledge used to encode the presented information varies greatly from expert to expert. Similarly, the details of the acquired cognitive structures (retrieval structures) to store information in retrieval form in long-term memory also differ. Chase and Ericsson (1982; Ericsson, 1985), however, found three principles of skilled memory that described the general characteristics of essentially all memory experts who have been systematically studied.

Studies of Particular Aspects of Expert Performance. Up to this point we have discussed studies of expert performance using tasks selected to capture the essence of that performance. It was pointed out that in many cases particular cognitive activities associated with expertise could be identified that could be more effectively examined in tasks designed to focus on those particular cognitive activities. For example, in their study of experts in physics, Chi et al. (1981) focused on the initial encoding of physics problems to account for these experts' immediate availability of plans for complete solutions to those problems. They asked experts and novices to sort a large number of physics problems into categories of similar problems. Consistent with the hypothesis that experts' encodings would incorporate information about solution methods, the experts' categories of problems reflected the physical principles underlying the solutions, whereas the novices' categories were based on the situations and objects mentioned in the problem text. In this case, the knowledge uncovered stands in close correspondence to the knowledge evoked during the solution of the physics problems. Several other investigators have used similar sorting methods to assess the immediate encodings of mathematical problems (Berger & Wilde, 1987, 1981), as well as encodings of pictures of situations in team sports (Allard & Starkes, 1991). It is, of course, possible to examine the knowledge of experts more generally. In their study of representation of expert knowledge, Olson and Biolsi (1991) discuss a wide range of methods. Attempts to measure knowledge about chess directly with psychometric tests have been quite successful, and scores on these tests show a clear correlation with rated chess performance (Charness, 1991).

During a study of the selection of the best move for an unfamiliar chess position, de Groot (1978) also found that the critical differences in cognitive processes relating to chess expertise occurred within the initial perception of the chess position. After a brief exposure to an unfamiliar chess position, the chess masters could give very informative verbal reports about the perceived characteristics of the presented chess position, along with virtually perfect recall of the locations of all chess pieces. In subsequent research, superior memory performance and superior perceptual performance of experts have been studied in specially designed tasks.

As reported earlier, Chase and Simon (1973) accounted for the superior memory performance of chess masters in terms of their storage of chess positions in short-term memory using complex independent chunks of chess pieces. The assumptions of storage in short-term memory and of independence of chunks have been seriously questioned by more recent investigators. Carefully designed studies of superior memory performance for chess positions, as reviewed by Charness (chapter 2, 1991), showed that chess experts store information about chess positions in long-term memory, not solely in short-term memory as Chase and Simon (1973) originally proposed.

Subsequent researchers have questioned Chase and Simon's (1973) assumption that chunks of chess pieces were distinct and that a given chess piece could therefore belong to only a single chunk. Chi (1978) showed that occasionally a chess piece can belong to more than one chunk, a finding that suggests relations between the chunks from a given chess position. On the basis of retrospective verbal reports of grand masters and masters after brief exposures to

chess positions, de Groot (1978) found clear evidence of perception of chess pieces in chunks, or complexes, as well as of encodings relating chunks to one another to form a global encoding of the position. It appears necessary to assume that global and integrating encodings account for the ability of chess experts to recall accurately more than one briefly presented chess position at a single trial (Frey & Adesman, 1976).

In analyses of superior memory performance in domains other than chess, evidence of global integration of the presented information has also been found (Egan & Schwartz, 1979; Reitman, 1976). Studies in other domains, however, have also revealed differences from the findings regarding chess experts. In domains with complex stimuli, such as medicine (Patel & Groen, 1991) and computer programming (Adelson, 1984), it is clear that part of the integration of the presented information involves identification of the relevant and critical information, and any analysis of subsequent recall must distinguish between relevant and irrelevant information. For different domains of expertise, the processes of encoding presented information will be quite different, depending on the demands of the particular type of expertise (Allard & Starkes, 1991). Expert dancers display superior memory for presented dance sequences, whereas skilled volleyball players can detect the location of the volleyball with superior speed. Superior perceptual processing has also been demonstrated as a function of chess expertise for tasks involving simple perceptual judgments about critical aspects of presented chess positions (Charness, 1991).

General Comments on the Analysis of Expert Performance. Once the expert performance can be elicited by a collection of tasks in the laboratory, the full range of methods in cognitive science can be applied to assess the mediating cognitive structures and processes. The mediating mechanism for an expert performance should be stable and not much influenced by the additional experience in the laboratory, as the laboratory experience will constitute only a minor fraction of the experts' total experience of tasks in their domains. In fact, an absence of further improvement during extended laboratory testing should provide a nice index for evaluating our ability to capture the mechanisms underlying the real-life expertise.

On the basis of this argument, one immediately realizes some potential dangers of studying aspects of "real" expert performance with tasks not encountered in the normal environments of the experts. If we provide an expert with unfamiliar tasks, we need to consider the possibility that the expert may resort to nonoptimal and unstable strategies that can be rapidly improved even during just a couple of sessions. With respect to memory for briefly presented chess positions, Ericsson and Oliver (Ericsson & Staszewski, 1989) found substantial improvement in the memory performance of a candidate chess master during a few months of testing. They found no evidence of changes in the mediating processes, however, only a marked speedup of the processes.

We have been unable to find much evidence concerning the effects of extended testing of experts. Ericsson (1985) reported several instances of marked improvements in the performance of memory experts when they were observed on several test occasions. In several cases the tests were separated by several years, and one cannot distinguish between the effects of testing and the

improvement due to accumulated experience outside the laboratory. Ericsson and Polson (1988b) found continual improvements in their expert waiter's performance of their standard task during about two years of weekly testing. It is likely that part of the observed speedup resulted from the particular constraints of the dinner orders studied. A more important determinant of the speedup, however, was the fact that the real-life task of memorizing dinner orders was not constrained by speed, because the customers required more time to decide on their dinner selections than the waiter needed to memorize them. Only in the laboratory situation with preselected dinner orders did the time required for memorization become critical.

In sum, differences between real-life situations and analogous laboratory tasks with respect to demands for maximum speed and the presented perceptual information are likely to lead to practice effects, even for experts, during extended testing. But as long as the practice effects for the experts remain comparatively small and the performance of the experts remains reliably superior to those for novices even after extended practice, we would claim that such a collection of tasks can successfully capture the superior expert performance.

The effects of extended practice for novices will provide a major source of empirical evidence as we now turn to a review of theoretical accounts of how the superior performance of experts can be acquired through extensive training.

Accounting for Superior Performance by Experts: The Third Step

In all the studies discussed earlier, the assessed mechanisms mediating superior performance implicated cognitive structures that were specific to the relevant task domains. The nature of the mediating cognitions allows us to infer that they reflect acquired knowledge and previous experiences in the domain. In order to account for those aspects of superior performance that are acquired, it is critical to understand the role of knowledge acquisition and the important effects of practice and training for their acquisition.

When we restrict ourselves to those task domains in which superior performance has been adequately captured, the empirical findings can be summarized relatively easily. The superior performance consists of faster response times for the tasks in the domain, such as the superior speed of expert typists, pianists, and Morse code operators. In addition, chess experts exhibit superior ability to plan ahead while selecting a move (Charness, 1981). In a wide range of task domains, experts have been found to exhibit superior memory performance.

What is acquired by experts? Superior performance in different domains reflects processes and knowledge specific to the particular domain. The challenge is to account for the widest range of empirical phenomena with the smallest of learning mechanisms and processes responsible for changes as a function of long-term practice. Because it is not possible to observe subjects during a decade of intensive practice, most of the empirical evidence is based on extrapolation of changes in performance found as a result of practice at laboratory tasks over much shorter terms. Another important constraint is that the proposed descriptions cannot posit performance capacities that would violate the known limits of human information processing.

In this section we shall consider various accounts concerning the processes and knowledge that experts have acquired. We shall first briefly describe the

Chase and Simon theory of expertise. Then we shall briefly review some of the empirical evidence concerning speedup of performance, superior memory performance, and superior ability to plan, with the intent of pointing to issues requiring further attention and elaboration.

The Chase and Simon Theory. Chase and Simon (1973) argued that the main differences among masters, experts, and novices in a wide range of domains were related to their immediate access to relevant knowledge. Chase and Simon's (1973) elegant theoretical account of chess expertise provided an account of how the masters rapidly retrieved the best move possibilities from long-term memory. The recognized configurations of chess pieces (chunks) served as cues to elicit the best move possibilities, which had been stored in memory at an earlier time. The chess masters' richer vocabulary of chunks thus played a critical role in the storage and retrieval of superior chess moves.

Within the same theoretical framework, the speedup in selecting moves can be accounted for in terms of recognition of chess configurations and direct retrieval of knowledge about appropriate move selections. Similarly, Chase and Simon (1973) proposed that the superior memory performance for the briefly presented chess positions was due to recognition of familiar configurations of chess pieces by the masters. The near-perfect recall by the chess masters, involving more than twenty chess pieces, was assumed to be mediated by approximately seven chunks or configurations—within the postulated limits of short-term memory.

Finally, with respect to planning, Chase and Simon (1973) outlined a mechanism whereby the experts' chess knowledge could be accessed in response to internally planned moves in the mind's eye. Given that no evidence was available to show that the depth of planning increased with a rise in the level of expertise (Charness, 1981), they did not consider the acquisition of such a mechanism.

Accounts Focusing on Practice and Learning. Across a wide range of tasks, an improvement in performance is a direct function of the amount of practice, and this relation can be remarkably accurately described by a power function (Newell & Rosenbloom, 1981). This consistent relation between performance and practice has been given a theoretical account by Newell and Rosenbloom (1981) using a uniform mechanism of learning chunks, which they explicitly relate to Chase and Simon's (1973) analysis of chess expertise.

It is possible to describe skill acquisition in a broader range of tasks and domains in which the subject at the outset does not have the prerequisite knowledge to produce error-free performance. In systematizing a large body of data on the acquisition of skills, Fitts (1964) proposed three different acquisition stages: The "cognitive stage" is characterized by an effort to understand the task and its demands and to learn to what information one must attend. The "associative stage" involves making the cognitive processes efficient to allow rapid retrieval and perception of required information. During the "autonomous stage," performance is automatic, and conscious cognition is minimal. More recently, Anderson (1982) provided a theoretical model with three different learning mechanisms, each corresponding to a stage of the Fitts model.

Anderson was able to derive a power law for relating performance to the amount of practice.

It is clear that the learning mechanisms that mediate increasing improvements from repeated practice trials must play important roles in the acquisition of expertise. It may even be useful to consider such mechanisms with an eye to identifying some limits to their applicability.

First, it is important to distinguish between practice and mere exposure or experience. It is well known that learning requires feedback in order to be effective. Hence, in environments with poor or even delayed feedback, learning may be slow or even nonexistent. Making predictions and forecasts for complex environments that are dynamically changing can present difficult information-extraction problems, which may, at least in part, account for the poor performance of expert consultants and decision-makers (Camerer & Johnson, 1991). In addition, merely performing a task does not ensure that subsequent performance will be improved. From everyday experience, anyone can cite countless examples of individuals whose performance never appears to improve in spite of more than 10 years of daily activity at a task. These observations deserve to be considered in more detail, but we shall limit ourselves to one issue relevant to research on expertise: On the basis of the foregoing considerations, one should be particularly careful about accepting one's number of years of experience as an accurate measure of one's level of expertise.

Second, the learning mechanisms discussed can account only for making the initial cognitive processes more efficient and ultimately automatic. In real-life perceptual motor skills, there exist a wide range of motor movements that can allow realization of a given goal. There is good evidence from sports that the beginner's spontaneously adopted baseline strokes in tennis or basic strokes in swimming are nonoptimal and that it is impossible to improve their efficiency by iterative refinement. Hence, the first thing a coach will do when beginners start training is to have them relearn their basic strokes to achieve correct form. Only then can the basic motor patterns be perfected through further training. It is thus possible that the final performance levels may reflect differences in the initial representations used by different subjects.

Third, once we are willing to consider the effects that result from weeks, months, and years of daily practice, it is likely that we cannot limit the consideration to purely cognitive effects on the central nervous system. Research on sports performance shows that extensive and intensive training is associated with a full range of changes related to the blood supply and the efficiency of muscles (Ericsson, 1990). Such changes will influence the speed of performance. It is possible that the correlations concerning speed of movements, as measured by maximum rate of tapping and speed of typewriting (Keele & Hawkins, 1982; Salthouse, 1984), should be considered not only as reflections of inherited characteristics but also as adaptations of the motor system during years of practice.

Finally, and most important, these types of learning mechanisms focus only on how performance can be made faster and more efficient; they do not take into account the acquisition of new cognitive structures, processes that are prerequisites for the unique ability of experts to plan and reason about problem situations.

Accounts Focusing on Memory Functioning. The Chase-Simon hypothesis that the superior memory of the expert reflects storage of more complex independent chunks in short-term memory has been seriously questioned, and most of the empirical evidence also suggests storage of interrelated information in long-term memory, as mentioned earlier. Even without the constraints of independence of chunks and storage in a limited-capacity short-term memory, human information-processing theory suggests a number of limits and processing constraints that must be taken into consideration in any acceptable account. But let us first review some of the empirical characteristics of the superior memory of experts.

Over a broad range of domains, experts have superior memory restricted to information in their domains of expertise. Furthermore, de Groot (1978) and Chase and Simon (1973) found that chess skill among a small number of subjects was monotonically related to their memory performance, which would suggest a high correlation between skill level and memory performance. Subsequent studies with representative samples involving large numbers of subjects found reliable correlations, but the strength of the association was lower than would have been expected from the Chase-Simon theory (Charness, 1991; Holding, 1985).

Although experts with decades of experience nearly always exhibit memory performance superior to that of subjects lacking expertise, there is at least one intriguing counterexample: Even though experts in mental calculation show far better memory performance for numbers than do normal subjects, their performance is far inferior to that of subjects who have practiced memorizing digits over extended periods (Chase & Ericsson, 1982; Ericsson, 1985). Whereas the mental-calculation experts rely predominantly on their vast mathematical knowledge of numbers, the trained subjects draw on a variety of knowledge essentially unrelated to mathematics. The most important difference between mental calculators and memory experts is that mental calculators require years and decades of practice to achieve memory performance comparable to what can be achieved by normal subjects after 50–100 hours of practice in a memory task. Hence, it is possible that the superior memory performance of experts has only a weak association with their expert knowledge.

Similarly, superior memory for briefly presented chess positions can be trained. Ericsson and Harris (1989) found that after 50 hours of practice, a subject without chess-playing experience was able to recall chess positions at a level of accuracy approaching that of some chess masters. In similarity to the digit-span experts, a close examination of the mediating processes revealed that the subject's performance was mediated by perceptually salient configurations of chess pieces, without implications for playing chess. Hence, it appears that by means of practice directed toward improving memory of performance, subjects without expertise can, after a couple of months of daily practice, match or surpass the superior memory performance of experts.

To account for the results concerning memory experts and long-term training studies, Chase and Ericsson (1981, 1982; Ericsson, 1985, 1988; Ericsson & Staszewski, 1989) proposed a skilled-memory theory to account for how memory performance can be improved within the known limits of human information processing. Chase and Ericsson proposed that experts can develop skilled

memory to rapidly store and retrieve information using long-term memory for information in their domains of expertise. Building on the distinction between a limited short-term memory and a vast long-term memory, this theory sees the key problem to be selective access to information stored in long-term memory. Skilled-memory theory postulates that at the time of encoding, experts acquire a set of retrieval cues that are associated in a meaningful way with the information to be stored. At a later time, the desired information can be retrieved from long-term memory by using the appropriate retrieval cue. After extensive practice using a stable set of retrieval cues with meaningful information in the domain, one's speed of encoding and retrieval is assumed to approach that for short-term memory. The best empirical evidence regarding the structure and operation of skilled memory comes from studies of subjects who achieved exceptional levels of performance on the digit-span task (Chase & Ericsson, 1981, 1982; Staszewski, 1987). The retrieval cues used for rapid storage of meaningful encodings of three- and four-digit groups (up to a total of more than a hundred digits) can be used to access digits in presented matrices in a manner earlier believed to require a raw visual image (Ericsson & Chase, 1982). Studies of other types of expertise have given clear evidence for retrieval cues indexing content (e.g., specific intermediate products in mental calculation) (Ericsson & Staszewski, 1989; Staszewski, 1988).

The most direct evidence suggesting the use of retrieval structures in chess comes from a series of studies with a candidate chess master by Ericsson and Oliver (Ericsson & Staszewski, 1989). They found that the chess master could read the description of the sequence of chess moves in a game and mentally generate the sequence of intermediate chess positions almost as fast as he could play out similar chess games by actually moving the pieces on a chessboard. During the process of mentally playing out the chess games, sometimes they would interrupt him and test his ability to name the piece on a given square for the current chess position, which he could do within a few seconds. In other experiments, his speed of access to different types of information for a briefly presented middle-game chess position was examined. The chess master could name the piece located on a given square within a second, and within seconds he could report the number of his opponent's pieces that were attacking a given square, which suggests remarkable availability of many different types of information about the presented chess position. Ericsson and Oliver (Ericsson & Staszewski, 1989) found evidence for rapid and flexible retrieval using a retrieval structure. This research raises the possibility that acquisition of expert-level chess skill involves the development of skilled memory for chess positions.

Once it is accepted that mediating mechanisms are acquired, that raises a number of challenging issues. One can no longer assume that superior performance is automatically achieved merely as a function of practice. The history of expert memory performance provides a number of cases in which individuals who have had extensive practice and experience have settled for suboptimal methods. Crutcher and Ericsson (Ericsson & Polson, 1988b) found that several waiters and waitresses who on a daily basis memorized dinner orders relied on less effective encoding methods than did the expert waiter JC, who exhibited vastly superior performance. Chase and Eriesson (1981, 1982) documented

extended problem-solving efforts by digit-span experts to identify strategies and encoding methods to increase their digit-span performance, as well as similar efforts by other subjects, whose performance never improved or did not improve beyond a certain level. When that evidence is considered together with studies of other memory experts (Ericsson, 1985, 1988) past and present, it appears that all memory experts rely on the same limited set of mechanisms (Chase & Ericsson, 1982). Given that most memory experts have not been instructed but have themselves discovered the structures necessary for their memory skills over extended periods, the importance of problem solving for their ultimate performance can hardly be overestimated. Similarly, studies of the development of a number of perceptual motor skills suggest the importance of discovered methods and strategies for performing tasks such as juggling (Norman, 1976). There appears to exist a wealth of phenomena such that successful performance in the future cannot be predicted on the basis of current performance. Similarly, there is no reason to believe that such problem solving is limited to the early stages in the development of expert performance.

Accounts Focusing on the Ability to Plan and Reason. Analyses in several different domains of expertise have revealed that experts engage in a number of complex mental activities involving reasoning that relies on mental models and internal representations. The most frequently studied activity has been the planning of chess moves. Charness (1981) found that the depth to which a possible move sequence for a chess position was explored was closely related to the level of chess skill, at least for chess players at or below the level of chess experts. Mental planning and evaluation of possible move sequences place greater demands on memory as the depth increases, and such a cognitive activity will be particularly tractable using acquired skilled memory to represent chess positions.

As noted earlier, de Groot (1978) found no reliable differences in regard to depth of search among advanced chess players with differing levels of chess ability. Holding (1985) suggested that the differences were too small to be detected, because of the small number of subjects. Charness (1989), however, presented a case study suggesting that the depth of search may increase with chess skill only up to some level of chess skill and then level off. One should also keep in mind that the task of searching for a move for a middle-game chess position is not designed to measure the capacity to make deep searches and hence may well reflect pragmatic criteria for sufficient depth of exploration to evaluate a prospective move.

In support of the findings of remarkable capacities to explore chess positions mentally, it is well known that chess players at the master level can play while blindfolded with only a minor reduction in chess capability without any prior specialized practice (Holding, 1985). In the absence of a strict time constraint, there appears to be no clear limit to the depth to which a chess master can explore a position. Ericsson and Oliver (Ericsson & Staszewski, 1989) found that a candidate chess master was able to access all the information about a mentally generated chess position rapidly and accurately, and they showed that the memory representation of the chess position was consistent with the characteristics of skilled-memory theory (Chase & Ericsson, 1982; Ericsson & Staszewski, 1989).

The need to represent and integrate large amounts of presented information internally is common to a wide range of different types of expertise. Charness (1989) showed that expertise at the game of bridge was closely linked with the capacity to generate successful plans for playing the cards in the optimum order. In medical diagnosis, the medical expert has to integrate many different pieces of information that are not simultaneously available perceptually. The internal representation of the presented medical information must be sufficiently precise to allow extensive reasoning and evaluation of consistency, but also must be sufficiently flexible to allow reinterpretation as new information becomes available (Lesgold et al., 1985; Patel & Groen, 1991). Anzai (1991) reviews the critical role of effective representations in solving physics problems and how methods of generating such representations can be developed through practice. In order to account for expertise, it is essential to describe emerging skills for managing extended memory demands, as well as their efficient processing and manipulation.

Comments on the Problem of Accounting for Expert Performance. Chase and Simon (1973) may have been correct in their claim that access to aggregated past experience is the single most important factor accounting for the development of expertise. More recent research, however, shows that to describe the structure of expertise accurately, several other factors must be considered, ranging from acquired skill allowing for an extended working memory to increased physiological efficiency of the motor system due to adaptation to intensive practice. We believe that the research on superior expert performance is benefited more by the development of a taxonomy of different types of mechanisms acquired through different types of learning and adaptation processes than by restricting the definition of expertise to a specific type of acquisition through learning.

Summary and Conclusion

In this chapter we initially contrasted the study of expertise with a number of other approaches studying outstanding and superior performance, and we found that one distinguishing feature was the claim that the superior performance was predominantly acquired. Drawing on the pioneering work on chess, we identified three important steps in the study of expertise: first, identification of a collection of representative tasks by means of which the superior performance of experts can be reproduced; second, analysis of the cognitive processes mediating that performance, followed by design of experimental tasks to elicit the critical aspects of such performance in a purer form; third, theoretical and empirical accounts of how the identified mechanisms can be acquired through training and practice.

The most effective approach to organizing the results across different domains of expertise is to propose a small number of learning mechanisms that can account for the development of similar performance characteristics in different domains within the limits of human information capabilities. There is now overwhelming empirical support for the theory of acquisition of skill with mechanisms akin to those originally proposed by Chase and Simon (1973). They

proposed their account as "simply a rough first approximation" (p. 252), and it would therefore make sense to seek a fuller account, both looking for the conditions limiting those principles and supplying other principles that can account for the complete range of performance capacities. Next we looked at some of those additional mechanisms. It would seem that one of the strengths of a generalized study of superior performances lies in a careful consideration of learning mechanisms and associated acquired characteristics uncovered across different domains.

We believe that both the excellent prospects and the clear-cut limitations of the expertise approach lie in its exacting methodological criteria, particularly the criterion that superior performance should be demonstrated as well as captured by a collection of laboratory tasks. To the extent that we are studying mechanisms and phenomena that have emerged as a result of intensive preparation during years or decades, we can be certain that tens or hundreds of hours of laboratory testing are not likely to alter their structure seriously. This affords excellent opportunities to examine and to describe carefully the mechanisms mediating the observed superior performance. In this regard, the superior expert performance is a phenomenon that is particularly well suited for laboratory study and experimental analysis.

A major limitation of the approach is the fact that many types of expertise have not yet been adequately captured. In some cases, the lack of success in capturing the essence of an expertise is so well documented that there may not be a legitimate phenomenon to study. Perhaps the most important limitation concerns the difficulty of studying the development of superior performance in real-life expertise. To understand the many factors underlying why some individuals attain the highest levels of performance whereas others do not, we need to broaden our approach. Indeed, in many cases we may well be forced to rely on correlational methods. As our ability to describe the structures of different types of expert performance improves, we shall be able to focus on the essential aspects, which can be monitored in longitudinal studies.

On the most general level, the study of expert performance provides us with a range of capacities and associated characteristics that can be acquired. A careful systematization of those should allow us to map out the potential for human performance that can be acquired through experience.

Acknowledgments

The thoughtful suggestions and comments on earlier drafts of this chapter by Ralf Krampe, Natalie Sachs-Ericsson, Herbert Simon, and Clemens Tesch-Römer are gratefully acknowledged.

References

Adelson, B. (1984). When novices surpass experts: The difficulty of the task may increase with expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 483–495.

Albert, R. S. (1983). Family positions and the attainment of eminence. In R. S. Albert (Ed.), *Genius and eminence* (pp. 141–154). Oxford: Pergamon Press. (Original work published 1980.)

Allard, F. & Burnett, N. (1985). Skill in sport. *Canadian Journal of Psychology*, *39*, 294–312.

Allard, F., & Starkes, J. L. (1991). Motor-skill experts in sports, dance, and other domains. In K. A. Ericsson & J. Smith (Eds.), *Toward a General Theory of Expertise* (pp. 126–152). New York: Cambridge University Press.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369–406.

Anzai, Y. (1991). Learning and use of representations for physics expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a General Theory of Expertise* (pp. 64–92). New York: Cambridge University Press.

Bachem, A. (1937). Various types of absolute pitch. *Journal of the Acoustical Society of America*, *9*, 146–151.

Baron, J. (1978). Intelligence and general strategies. In G. Underwood (Ed.), *Strategies in information processing* (pp. 403–450). London: Academic Press.

Bennett, H. L. (1983). Remembering drink orders: The memory skill of cocktail waitresses. *Human Learning*, *2*, 157–169.

Berger, D. E., & Wilde, J. M. (1987). A task analysis of algebra word problems. In D. E. Berger, K. Pezdek, & W. P. Banks (Eds.), *Application of cognitive psychology: Problem solving, education and computing* (pp. 123–137). Hillsdale, NJ: Erlbaum.

Bloom, B. S. (Ed.). (1985). *Developing talent in young people*. New York: Ballantine Books.

Brady, P. T. (1970). The genesis of absolute pitch. *Journal of the Acoustical Society of America*, *48*, 883–887.

Camerer, C. F., & Johnson, E. J. (1991). In K. A. Ericsson & J. Smith (Eds.), *Toward a General Theory of Expertise* (pp. 195–217). New York: Cambridge University Press.

Carroll, J. B. (1978). How shall we study individual differences in cognitive abilities? Methodological and theoretical perspectives. *Intelligence*, *2*, 87–115.

Cattell, R. B. (1963). The personality and motivation of the researcher from measurements of contemporaries and from bibliography. In C. W. Taylor & F. Barron (Eds.), *Scientific creativity: Its recognition and development* (pp. 119–131). New York: Wiley.

Cattell, R. B., & Drevdahl, J. E. (1955). A comparison of the personality profile (16 PF) of eminent researchers with that of eminent teachers and administrators, and of the general population. *British Journal of Psychology*, *46*, 248–261.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 641–653.

Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology*, *33*, 1–6.

Charness, N. (1981). Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 467–476.

Charness, N. (1989). Expertise in chess and bridge. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 183–208). Hillsdale, NJ: Erlbaum.

Charness, N. (1991). Expertise in chess: the balance between knowledge and search. In K. A. Ericsson & J. Smith (Eds.), *Toward a General Theory of Expertise* (pp. 39–63). New York: Cambridge University Press.

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141–189). Hillsdale, NJ: Erlbaum.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, pp. 1–58). New York: Academic Press.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Erlbaum.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.

Chi, M. T. H., Glaser, R. & Rees, E. (1982). Expertise in problem solving. In R. S. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 1–75). Hillsdale, NJ: Erlbaum.

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*, 257–273.

Cooper, L. A., & Regan, D. T. (1982). Attention, perception and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 123–169). Cambridge: Cambridge University Press.

de Groot, A. (1978). *Thought and choice in chess*. The Hague: Mouton. (Original work published 1946.)

Doll, J., & Mayr, U. (1987). Intelligenz und Schachleistung—eine Untersuchung an Schachexperten. [Intelligence and achievement in chess—A study of chess masters]. *Psychologische Beiträge*, *29*, 270–289.

Dörner, D., & Schölkopf, J. (1991). Controlling complex systems; or, Expertise as "grandmother's know-how." In K. A. Ericsson & J. Smith (Eds.), *Toward a General Theory of Expertise* (pp. 218–239). New York: Cambridge University Press.

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition*, *7*, 149–158.

Elo, A. E. (1978). *The rating of chessplayers, past and present*. London: Batsford.

Engle, R. W., & Bukstel, L. H. (1978). Memory processes among bridge players of differing expertise. *American Journal of Psychology*, *91*, 673–689.

Ericsson, K. A. (1985). Memory skill. *Canadian Journal of Psychology*, *39*, 188–231.

Ericsson, K. A. (1988). Analysis of memory performance in terms of memory skill. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 137–179). Hillsdale, NJ: Erlbaum.

Ericsson, K. A. (1990). Peak performance and age: An examination of peak performance in sports. In P. B. Baltes & M. M. Baltes (Eds.), *Successful aging: Perspectives from the behavioral sciences* (pp. 164–195). Cambridge: Cambridge University Press.

Ericsson, K. A., & Chase, W. G. (1982). Exceptional memory. *American Scientist*, *70*, 607–615.

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*, 1181–1182.

Ericsson, K. A., & Crutcher, R. J. (1990). The nature of exceptional performance. In P. B. Baltes, D. L. Featherman, & R. M. Lerner (Eds.), *Life-span development and behavior* (Vol. 10, pp. 187–217). Hillsdale, NJ: Erlbaum.

Ericsson, K. A., & Faivre, I. (1988). What's exceptional about exceptional abilities? In L. K. Obler & D. Fein (Eds.), *The exceptional brain: Neuropsychology of talent and special abilities* (pp. 436–473). New York: Guilford.

Ericsson, K. A., & Harris, M. (1989). *Acquiring expert memory performance without expert knowledge: A case study in the domain of chess*. Unpublished manuscript.

Ericsson, K. A., & Oliver, W. (1988). Methodology for laboratory research on thinking: Task selection, collection of observation and data analysis. In R. J. Sternberg & E. E. Smith (Eds.), *The psychology of human thought* (pp. 392–428). New York: Cambridge University Press.

Ericsson, K. A., & Polson, P. G. (1988a). An experimental analysis of a memory skill for dinner-orders. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 305–316.

Ericsson, K. A., & Polson, P. G. (1988b). Memory for restaurant orders. In M. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 23–70). Hillsdale, NJ: Erlbaum.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: Bradford Books/MIT Press.

Ericsson, K. A., & Staszewski, J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235–267). Hillsdale, NJ: Erlbaum.

Feltovich, P. J., Johnson, P. E., Moller, J. H., & Swanson, D. B. (1984). LCS: The role and development of medical knowledge in diagnostic expertise. In W. J. Clancey & E. H. Shortliffe (Eds.), *Readings in medical artificial intelligence* (pp. 275–319). Reading, MA: Addison-Wesley.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 243–285). New York: Academic Press.

Frey, P. W., & Adesman, P. (1976). Recall memory for visually presented chess positions. *Memory and Cognition*, *4*, 541–547.

Galton, F. (1869). *Hereditary genius*. New York: Macmillan.

Gilhooly, K. J., Wood, M., Kinnear, P. R., & Green, C. (1988). Skill in map reading and memory for maps. *Quarterly Journal of Experimental Psychology*, *40A*, 87–107.

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Hayes, J. R. (1981). *The complete problem solver*. Philadelphia: Franklin Institute Press.

Hinsley, D. A., Hayes, J. R., & Simon, H. A. (1977). From words to equations: Meaning and representation in algebra word problem. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 89–108). Hillsdale, NJ: Erlbaum.

Holding, D. H. (1985). *The psychology of chess skill*. Hillsdale, NJ: Erlbaum.

Howe, M. J. A., & Smith, J. (1988). Calendar calculating in "idiot savants": How do they do it? *British Journal of Psychology*, *79*, 371–386.

Hunt, E. (1980). Intelligence as an information processing concept. *Journal of British Psychology*, *71*, 449–474.

Jeffries, R., Turner, A. A., Polson, P. G., & Atwood, M. E. (1981). The processes involved in designing software. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 255–283). Hillsdale, NJ: Erlbaum.

Johnson, P. E., Duran, A. A., Hassebrock, F., Moller, J., Prietula, M., Feltovich, P. J., & Swanson, D. B. (1981). Expertise and error in diagnostic reasoning. *Cognitive Science*, *5*, 235–283.

Keele, S. W., & Hawkins, H. L. (1982). Explorations of individual differences relevant to high level skill. *Journal of Motor Behavior*, *14*, 3–23.

Kelley, H. P. (1964). Memory abilities: A factor analysis. *Psychometric Society Monographs*, *11*, 1–53.

Kliegl, R., Smith, J., & Baltes, P. B. (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. *Developmental Psychology*, *25*, 247–256.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*, 1335–1342.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1985). *Expertise in a complex skill: Diagnosing X-ray pictures*. LRDC, University of Pittsburgh Technical Report.

Lewis, C. (1981). Skill in algebra. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 85–110). Hillsdale, NJ: Erlbaum.

Luria, A. R. (1968). *The mind of a mnemonist*. New York: Avon.

McCurdy, H. G. (1983). The childhood pattern of genius. In R. S. Albert (Ed.), *Genius and eminence* (pp. 155–169). Oxford: Pergamon Press. (Original work published 1957.)

McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, *13*, 307–325.

Miller, G. A. (1956). The magical number seven, plus or minus two. *Psychological Review*, *63*, 81–97.

Morris, P. E., Gruneberg, M. M., Sykes, R. N., & Merrick, A. (1981). Football knowledge and the acquisition of new results. *British Journal of Psychology*, *72*, 479–483.

Morris, P. E., Tweedy, M., & Gruneberg, M. M. (1985). Interest, knowledge and the memorization of soccer scores. *British Journal of Psychology*, *76*, 415–425.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A. (1976). *Memory and attention* (2nd ed.). New York: Wiley.

Oakes, W. F. (1955). An experimental study of pitch naming and pitch discrimination reaction. *Journal of Genetic Psychology*, *86*, 237–259.

Oden, M. H. (1968). The fulfillment of promise: Forty-year follow-up of the Terman gifted group. *Genetic Psychology Monographs*, *77*, 3–93.

Olson, J. R., & Biolosi, K. J. (1991). Techniques for representing expert knowledge. In K. A. Ericsson & J. Smith (Eds.), *Toward a General Theory of Expertise* (pp. 240–285). New York: Cambridge University Press.

Patel, V. L., & Groen, G. L. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, *10*, 91–116.

Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: a critical look. In K. A. Ericsson & J. Smith (Eds.), *Toward a General Theory of Expertise* (pp. 93–125). New York: Cambridge University Press.

Reitman, J. (1976). Skilled perception in go: Deducing memory structures from interresponse times. *Cognitive Psychology*, *8*, 336–356.

Resnick, L. B. (Ed.). (1976). *The nature of intelligence*. Hillsdale, NJ: Erlbaum.

Roe, A. (1953). A psychological study of eminent psychologists and anthropologists, and a comparison with biological and physical scientists. *Psychological Monographs*, *67*, 1–55.

Saariluoma, P. (1984). *Coding problem spaces in chess: A psychological study*. Helsinki: Societas Scientiarum Fennica.

Salthouse, T. A. (1984). Effects of age and skill in typing. *Journal of Experimental Psychology: General*, *113*, 345–371.

Scardamalia, M., & Bereiter, C. (1991). Literate expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a General Theory of Expertise* (pp. 172–194). New York: Cambridge University Press.

Silver, E. A. (1981). Recall of mathematical information: Solving related problems. *Journal for Research and Mathematical Education*, *12*, 54–64.

Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 325–348). Hillsdale, NJ: Erlbaum.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, *61*, 394–403.

Simon, H. A., & Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology*, *8*, 165–190.

Sloboda, J. (1976). Visual perception of musical notation: Registering pitch symbols in memory. *Quarterly Journal of Experimental Psychology*, *28*, 1–16.

Sloboda, J. (1991) Musical expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a General Theory of Expertise* (pp. 153–171). New York: Cambridge University Press.

Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*, 275–290.

Stanley, J. C., George, W. C., & Solano, C. H. (1977). *The gifted and creative: A fifty-year perspective*. Baltimore: Johns Hopkins University Press.

Staszewski, J. J. (1987). *The psychological reality of retrieval structures: An investigation of expert knowledge*. Unpublished doctoral dissertation, Cornell University, Ithaca, NY.

Staszewski, J. J. (1988). Skilled memory and expert mental calculation. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 71–128). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (Ed.). (1982). *Handbook of human intelligence*. Cambridge University Press.

Terman, L. M., & Oden, M. H. (1947). *Genetic studies of genius. Vol. 4: The gifted child grows up*. Stanford, CA: Stanford University Press.

Thorndyke, P. W., & Stasz, C. (1980). Individual differences in procedures for knowledge acquisition from maps. *Cognitive Psychology*, *12*, 137–175.

Tyler, L. E. (1965). *The psychology of human differences*. New York: Appleton-Century-Crofts.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Varon, E. J. (1935). The development of Alfred Binet's psychology. *Psychological Monographs*, *46* (Whole No. 207).

Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem-solving skill in the social sciences. *Psychology of Learning and Motivation, 17*, 165–213.

Voss, J. F., Vesonder, G. T., & Spilich, G. J. (1980). Text generation and recall by high-knowledge and low-knowledge individuals. *Journal of Verbal Leaning and Verbal Behavior*, *19*, 651–667.

Wechsler, D. (1952). *The range of human capacities*. Baltimore: Williams & Wilkins.

Wilson, R. S. (1986). Twins: Genetic influence on growth. In R. M. Malina & C. Bouchard (Eds.), *Sports and human genetics* (pp. 1–21). Champaign, IL: Human Kinetics Publishing.

Chapter 23

Three Problems in Teaching General Skills

John R. Hayes

We need educational practices that will help people to adapt to a rapidly changing environment. We want students to acquire general skills—skills likely to transfer to the new situations that will face them. I was asked to consider whether there are any general skills to be taught. I believe that there are, and I also believe that it will not be as easy as we would like to teach them.

In this chapter, I discuss three problems that anyone who wants to teach general skills must face. The first is that proficiency in some general skills may require vast bodies of knowledge—knowledge that could take years to acquire. A second problem is that the task of teaching learning and thinking skills may be complicated by their number. If there were just three or five candidate strategies, it would be a relatively straightforward matter to set about evaluating them and teaching the useful ones. However, I argue that there are actually several hundred plausible strategies we might teach. Finally, the third problem with teaching general skills is that even after we identify a useful strategy and teach it successfully in one application, students may and frequently do fail to transfer that strategy to other applications.

The Requirements for Knowledge

The work of DeGroot (1965), Simon and Chase (1973), and Simon and Gilmartin (1973) has demonstrated clearly that skillful chess players employ an enormous amount of knowledge of chess patterns. To acquire this knowledge, the chess player must spend thousands of hours of preparation—playing chess, reading chess magazines, and studying chess positions. Simon and Chase (1973) note that it is very rare for a person to reach the grandmaster level of skill with less than 10 years of intensive study.

I do not want to argue that chess is an important general skill. It may well be that chess knowledge equips people to do little beyond playing chess. However, I do want to argue that there are valuable skills—specifically musical composition, painting, and perhaps other skills—that like chess depend on acquiring large bodies of knowledge. To explore this question in the area of music, I examined the lives of famous composers.

I started my investigation with the incredibly precocious Mozart because he is the composer who seems least likely to have required a long period of

From chapter 17 in *Thinking and Learning*, Vol. 2, ed. J. Segal, S. Chipman, and R. Glaser (Hillsdale, NJ: Erlbaum, 1985), 391–405. Reprinted with permission.

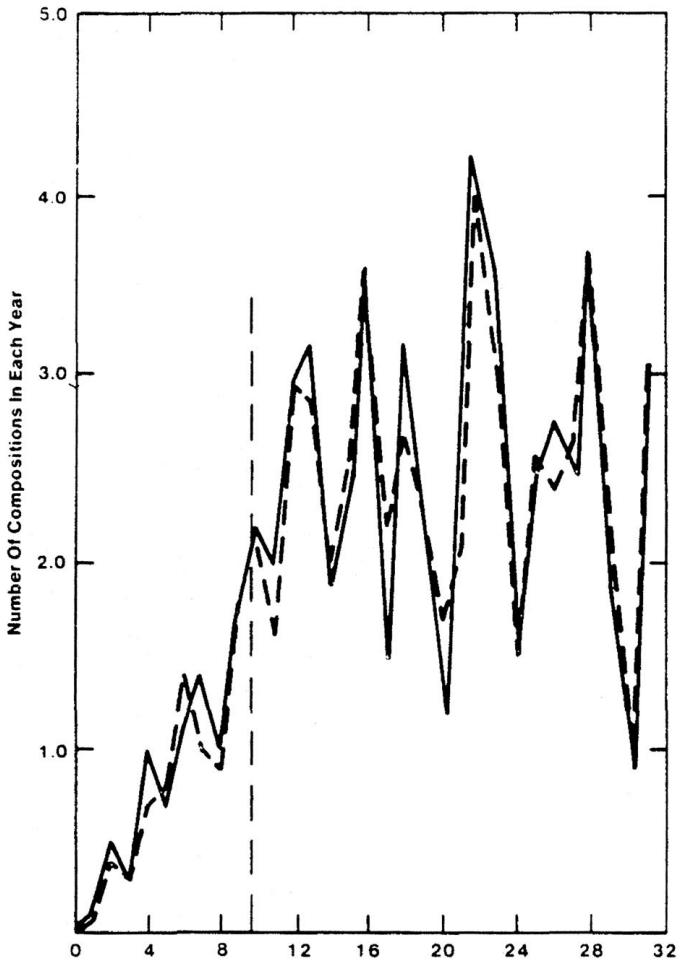


Figure 23.1
Graph of Mozart's compositions. The year marked 0 on the graph is 1760, the year when Mozart was 4 and began intensive musical training. The solid line in the figure is based on information from *Grove's Dictionary of Music* (1954). The dashed line is based on Koechel's listings (1965) as revised by modern musicologists. These two sources are in reasonable agreement about what works were produced when.

preparation. He began to study music at four and wrote his first symphony at the age of eight.

I have graphed the number of works that Mozart produced in each year of his career in figure 23.1. The figure shows that Mozart's productivity increased steadily for the first 10 or 12 years of his career, as reported by Groves (1954) and Koechel (1965). It also shows that Mozart did produce works in the very early part of his career when he had had only a year or two of preparation. If these are works of very high quality, then we could conclude, for Mozart at least, that long preparation is not a necessary condition for the production of outstanding musical works. However, these early works may not be of very high quality. Perhaps they have been preserved for their historical rather than for their musical value.

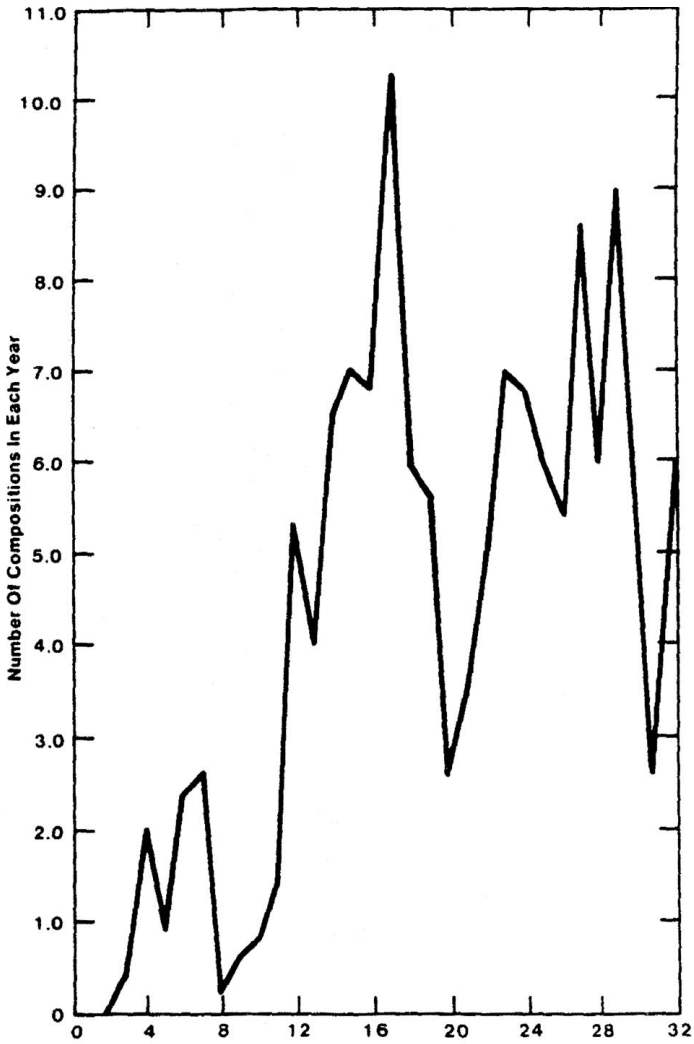


Figure 23.2
Number of recordings listed in *Schwann's Guide* (August, 1979) of works written in each year of Mozart's career.

To obtain some measure of the quality of Mozart's work, I turned to Schwann's *Record and Tape Guide*. I reasoned that an excellent work is likely to be recorded more often than a poor one. The decision to record a work presumably reflects both musical judgment and popular taste—that is, it reflects the musical judgment by a conductor that the work is worthwhile and the belief of the record companies that the record will sell.

Figure 23.2 shows the number of recordings listed in Schwann's guide (August, 1979) of works written in each year of Mozart's career. Although about 12% of Mozart's works were written in the first 10 years of his career, only 4.8% of the recordings came from this early period. Further, many of the recordings of early works are included in collections with labels such as, "The Complete

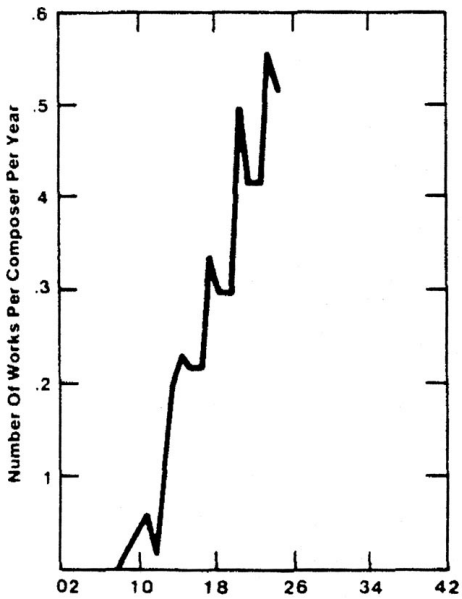


Figure 23.3
A graph of the careers of all composers in Hayes' study.

Symphonies of Mozart." Perhaps the early works were included for reasons of completeness rather than excellence. When recordings included in complete collections are omitted from the calculations, the percentage of recordings in this early period drops to 2.4. These observations suggest that Mozart's early works are not of the same high quality as his later ones. The music critic, Harold Schonberg (1970), is of the same opinion. He says:

> It is strange to say of a composer who started writing at six, and lived only thirty-six years, that he developed late, but that is the truth. Few of Mozart's early works, elegant as they are, have the personality, concentration, and richness that entered his music after 1781. (pp. 82–84)

In 1782, Mozart was in the 21st year of his career.

Some works are recorded two or three times in different complete collections. Therefore, to weed out works recorded for reasons other than musical quality, I defined a masterwork (for the purposes of this study) as one for which five different recordings are currently listed in Schwann's guide. By this definition, Mozart's first masterwork was written in the 12th year of his career.

To explore the question about creativity and preparation more generally, I searched for biographical material about all the composers discussed in Schonberg's *The Lives of the Great Composers* (1970). For 76 of these composers, I was able to determine when they started intensive study of music. Incidentally, all these composers had at least one work listed in Schwann's guide, and 64 had one or more works available on five different records.

In figure 23.3 all of the careers of the composers are shown on the same scale, that is, the 10th year of Handel's career is graphed in the same place as the 10th

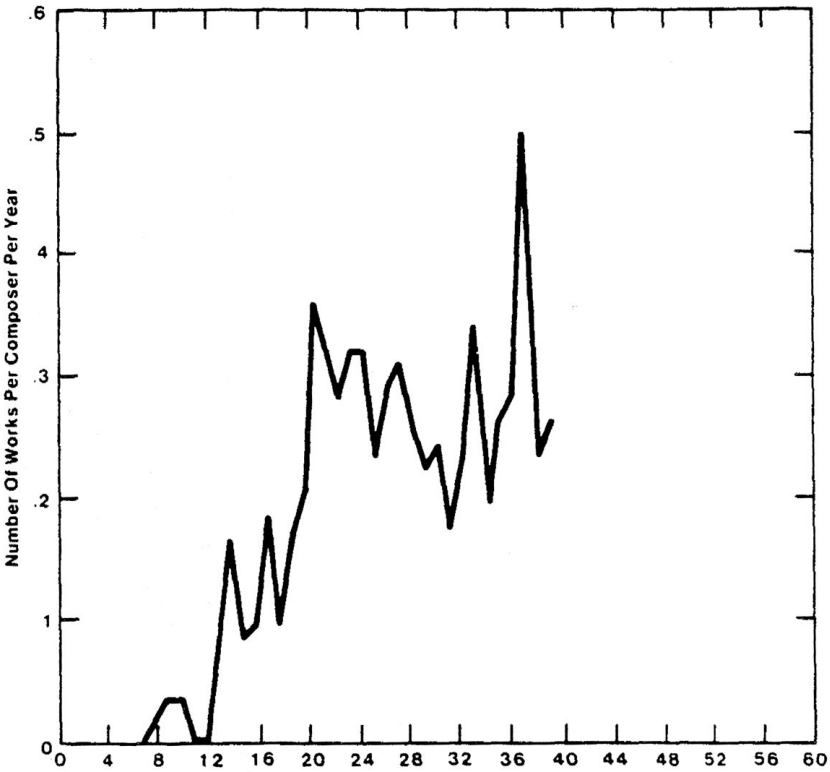


Figure 23.4
Graph showing that composers maintain their productivity through the 40th year of their career.

year of Brahms' career. The figure shows that very few composers produced masterworks with less than 10 years of preparation. There are just three exceptions: Satie's "Trois Gynopédies," written in year 8; Shostakovich's Symphony #1, and Paganini's Caprices, both written in year 9. Between year 10 and year 25, there is a rapid and essentially linear increase in productivity from almost zero to slightly more than half a work per composer per year.

I have not continued figure 23.3 beyond year 25 because to do so would have given a misleading impression of changes in productivity with age. All the composers in our sample had careers of 25 years or more. However, some composers died quite young. Schubert, for example, died in the 25th year of his career and Mozart died in the 31st year of his. Famous composers who die young tend to be unusually productive. This observation does not imply that especially creative musicians compose themselves to death. Rather, we believe that it is a statistical artifact captured by Hayes' maxim, "Late bloomers who want to be famous shouldn't die young."

If Handel and Verdi had died as young as Schubert, they would probably not be considered major composers. All their major works were written after they had been in music for 25 years. Averaging together short and long careers would make it appear that composers get less productive after 25 or 30. Actu-

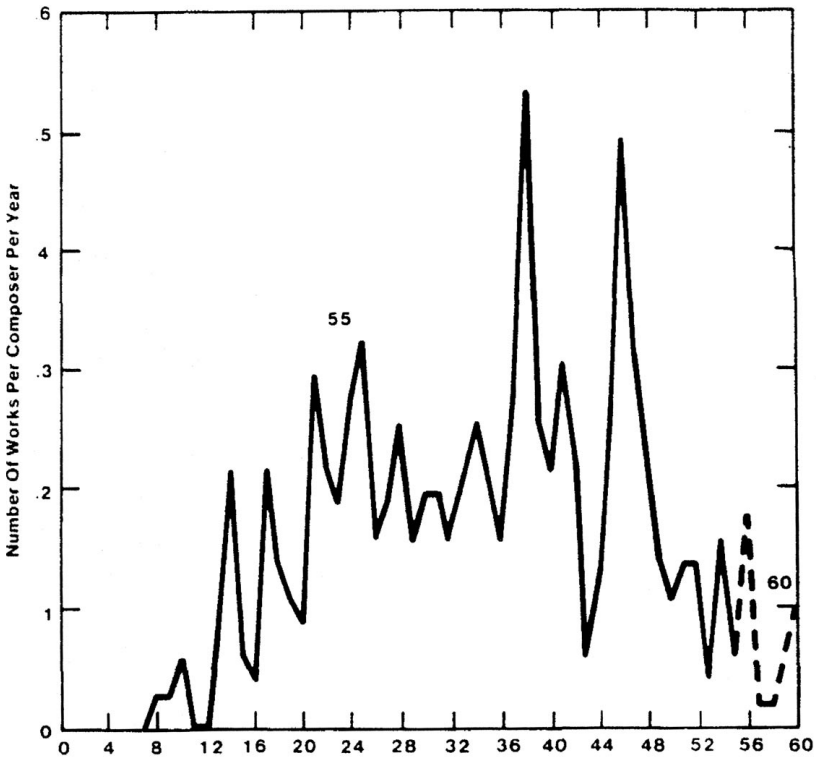


Figure 23.5
Graph indicating a decline in productivity for composers, beginning at about the 50th year of their careers.

ally, this is not so. This distortion is avoided in figure 23.4 by including only composers who have had careers of 40 years or more, and in figure 23.5 by including only composers who had careers of 55 and 60 years or more.

Figure 23.4 shows that composers maintain their productivity at least through the 40th year of their careers. Figure 23.5 indicates that a decline in productivity begins at about 50 years into the composers' careers. These figures, of course, do not take the composer's health into account. If we were to consider only composers in good physical and mental health, the decline in productivity might be much less marked. Clearly, productivity can continue far beyond the 50th year of the composer's musical career. For example, Albeniz's first masterwork was written in the 72nd year of his career!

It is reasonable to ask whether the important factor in the composers' productivity is really preparation or if perhaps the important factor is simply age. It is conceivable, for example, that composers have to be, say, 16 or 22, before they can write good music. Perhaps it is experience in life rather than experience in music that is critical. To test this possibility, I divided the composers into three groups. The first consisted of 14 composers who had begun their careers between the ages of 3 and 5. The second consisted of 30 composers who began their careers between 6 and 9 years of age. The third group consisted of 20 composers who began their careers at 10 or later.

I reasoned that if age were the critical factor, those who started their careers early would have to wait longer to produce good work than those composers who started late. In fact, this was not the case. The median number of years to first notable composition was 16.5 for the first group, 22 for the second group, and 21.5 for the third group.

It appears then that what composers need to write good music is not maturing but rather musical preparation. The results make it dramatically clear that no one composes outstanding music without first having about 10 years of intensive musical preparation.

These results *do not* mean that there is no such thing as genius. They *do not* mean that just anyone with 10 to 25 years of experience can write great music. They *do* mean that even a person endowed with the genius of Mozart or Beethoven will still need 10 years or more of intense preparation to realize his or her potential.

Do painters also require years of intense preparation to be productive? Sandra Bond, Carol Janik, Felicia Pratto, and I have conducted a parallel study of painters designed to answer this question. For the purpose of the study, we defined an outstanding painting as one reproduced in any of 11 standard histories of art. We defined the beginning of the artist's career as the point at which he or she began intensive study of art. For many, this point was marked by the beginning of an apprenticeship or by entry into an art academy.

Figure 23.6 shows how productivity (the number of outstanding works produced per year per painter) varies with the painters' years of experience in the profession. The 16-year curve presents data for 132 painters who had careers of at least 16 years. The 40-year curve presents data for 102 painters who had careers of at least 40 years.

The results for painters are generally similar to those for composers. The productivity curve for painters has an initial period of very low productivity followed by a period in which productivity increases very rapidly. Then there is a long period of stable productivity followed by a gradual decline. The period of rapid increase in productivity occurs between 6 and 12 years for painters rather than between years 10 and 24 as was observed for composers. This difference may reflect differences in the nature of the skills involved in the fields or differences in our criteria for identifying outstanding works in the two fields. In part, we believe it reflects a difference in the sensitivity of our biographical measures to experience in music and art. We believe that parents are more likely to notice and record musical activity, perhaps because it makes a noise, than drawing. For many of the painters, there was evidence of early but undated drawing activity. Because it was unquantifiable, this early experience could not be included in our study as part of the painter's preparation.

If skill in chess, musical composition, and painting depend on large amounts of knowledge, it is easy to believe that there are other skills that do so as well, for example, skills in writing poetry, fiction, or expository prose, and skill in science, history, and athletics as well as many others. Strategies may help in acquiring or executing such skills. However, it is unlikely that the use of strategies can circumvent the need to spend large amounts of time acquiring a knowledge base for such skills.

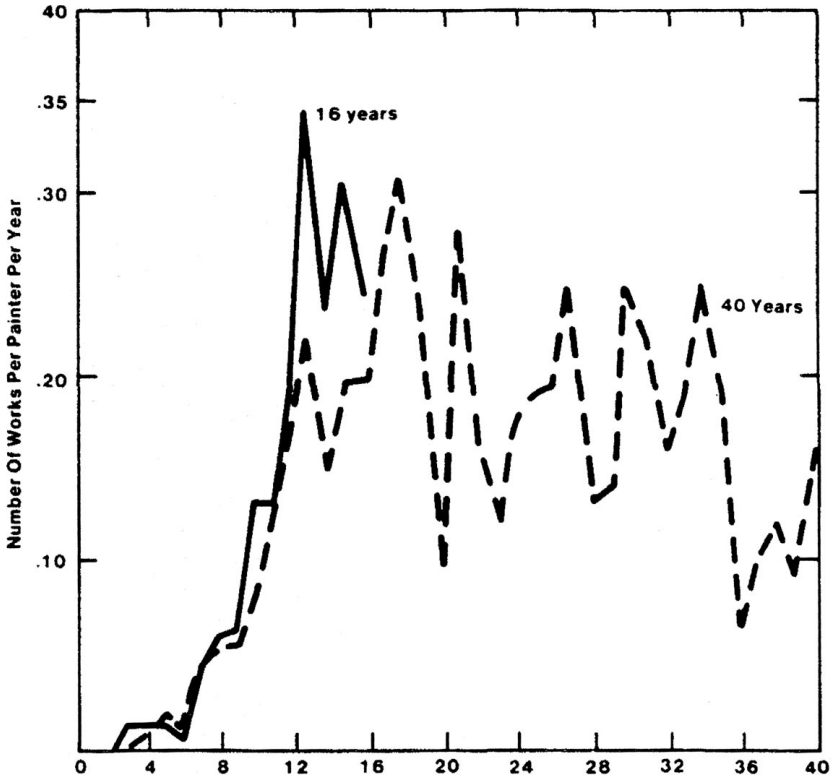


Figure 23.6
A graph of the careers of all the painters in Hayes' study.

The Large Number of Reasonable Strategies

People differ in their proficiency in learning, in reasoning, and in problem solving, and in the strategies they employ to do these things. It seems reasonable to teach the strategies used by good learners and thinkers to those who are less proficient. I teach a course at Carnegie Mellon University intended to do just this. It is a freshman course that assumes little sophistication on the part of the student. Its structure is reflected in my text, *The Complete Problem Solver* (1981). In the course, I teach basic strategies in problem finding, representation, solution search, decision making, memory, and learning. In examining the course materials, I was surprised to find that I present at least 50 different strategies during the semester. The strategies, listed in table 23.1, include such diverse procedures as searching for counterexamples, working backward, perspective drawing, brainstorming, fractionation, satisficing, the keyword method, and time management skills.

When I say that the strategies are diverse, I mean that they are quite distinct. They are not simple variants of a few general strategies. They have different purposes and different contexts and must be taught separately. I am not suggesting that the strategies taught in my course are exactly the right ones. Of

course, each of them seems plausible to me, but most of them have not been evaluated. I am suggesting, though, that the number of plausible strategies is large.

Another person teaching a course with the same orientation as mine might choose to teach many of the same strategies. However, there are many different ways to orient a basic strategies course. For example, a course could be focused on human relations problems or on math, on writing or on spoken communication, on learning through reading or on the analysis of arguments. Further, courses could be aimed at college students, or high school or grammar school students. Each focus and each age level would require a very different selection of strategies. Polya's *How To Solve It* (1973), which focuses on mathematics, includes some 60 strategies. Relatively few of these, perhaps 15, overlap with those in table 23.1. Taken together, these courses might easily include several hundred different plausible strategies—perhaps as many as a thousand.

The large number of plausible strategies poses a problem for us. Evaluating hundreds of strategies is a major research task—one that will not soon be completed. Fortunately, some excellent strategy evaluation work is already under way. However, until much more is done, choosing which strategies to teach will involve guesses and potentially faulty judgment.

Being mistaken about strategies can have serious consequences. For example, a student in my course had written an essay that had omitted an important qualification of its major point. The student's teaching assistant pointed out this flaw and precipitated the following dialogue:

Student: "I know, but I already have three paragraphs."

TA: "What?"

Student: "I've already proposed three ideas, so I've used up my three paragraphs."

TA: "What?"

Student: "An essay has just three paragraphs."

TA: "What?"

Student: "Beginning, middle, and end. So you see, I just couldn't add an extra idea."

Clearly, this student has learned some rather odd strategies for writing that put serious constraints on what he was able to do with language.

College English teachers report that they frequently observe equally bizarre strategies. One teacher, for example, reported that a student had asked her, "Aren't you going to give me extra credit because I didn't use any pronouns in my paper?"

Failure to Generalize Strategies

Sometimes a strategy that ought to generalize does not. Herb Simon and I have worked a good deal with problem isomorphs (1976)—that is, with sets of problems that have the same underlying structure, but different cover stories. For example, we have developed and studied a set of problems, all of which are identical in form to the famous Tower of Hanoi puzzle. Four of these problems,

Table 23.1
Strategies taught in problem-solving course

Problem Finding
Bug lists for identifying needed innovations
Search for counter-examples
Search for alternative interpretations
Representation
When in difficulty, examine problem statement to see if information has been properly extracted
When in difficulty, search for a new problem representation
Change point of view
Choose new sensory code, e.g., imagery
Work backwards
Try hypothetical reasoning
Try proof by contradiction
Be active in defining ill-defined problems by
Making gap filling decisons
Trying to solve the problem as a method for understanding it
Use external representations where possible
Use perspective drawing
Use matrices for keeping track of information
Use drawing to find implicit relations in the problem
Search
Brainstorm
Use heuristic search where possible
Planning
Means-End analysis
Auxiliary problems
Fractionation
Analogies
Decision Making
Explicit decision methods help
Satisficing
Dominance
Additive weighting
Expected value
Signal detection model
Bayes' Theorem
Minimax
Minimize maximum regret
Cost benefit analysis
Bargaining strategies
Schelling's task
Memory and Learning
Use external memory aids
Mnemonics
Method of loci
Keyword method

Table 23.1
(continued)

Learning strategies
Elaborative rehearsal
Notice hierarchical structure
Use overlearning
Monitor own learning
Generate examples
Use information in word roots
Evaluation
Check results
Get external criticism
General
Consolidate
Examine own process
Time management skills

which involve the actions of an imaginary set of "monsters," are shown in table 23.2. In the first puzzle, the monsters pass globes of various sizes back and forth; in the second, they move themselves from globe to globe; in the third, they change the sizes of the globe; and in the fourth, they change their own sizes.

Ideally, because these problems are formally identical, people who have solved one of them should behave as if they had solved them all. In fact, this is not the case. There is a lot of transfer between problems that involve moving either monsters or globes, and there is a lot of transfer between problems that involve changing the sizes of either monsters or globes. But there is relatively little transfer between move and change problems.

Failure of transfer is a frustrating reality in our classrooms. A statistics teacher at CMU who had taught the Poisson distribution to this class through a distance example was surprised that the next day his students could not apply the distribution to an example involving time.

Possible Responses to These Problems

The possibility that mastery of a field may take many years is an important item of metacognitive knowledge that we ought to teach to our students. Some students may be inappropriately discouraged by early setbacks because they believe that failure indicates lack of talent rather than lack of knowledge. Others, perhaps too well endowed with self-confidence, may believe that they are destined to perform great acts of creativity with little or no effort on their part. Some may even defend themselves against knowledge on the grounds that it may spoil the purity of their individual spark. Students of either type could profit by learning that large quantities of knowledge may be essential for skilled performance in their fields.

Table 23.2
Four monster problems

1. *Monster Problem (Transfer Form 1)*

   Three five-handed extraterrestrial monsters were holding three crystal globes. Because of the quantum-mechanical peculiarities of their neighborhood, both monsters and globes come in exactly three sizes with no others permitted: small, medium, and large. The medium-sized monster was holding the small globe; the small monster was holding the large glove; and the large monster was holding the medium-sized globe. Because this situation offended their keenly developed sense of symmetry, they proceeded to transfer globes from one monster to another so that each monster would have a globe proportionate to its own size.

   Monster etiquette complicated the solution of the problem because it requires that:

   1. only one globe may be transferred at a time;
   2. if a monster is holding two globes, only the larger of the two may be transferred;
   3. a globe may not be transferred to a monster who is holding a larger globe.

   By what sequence of transfers could the monsters have solved this problem?

2. *Monster Problem (Transfer Form 2)*

   Three five-handed extraterrestrial monsters were standing on three crystal globes. Because of the quantum-mechanical peculiarities of their neighborhood, both monsters and globes come in exactly three sizes with no others permitted: small, medium, and large. The medium-sized monster was standing on the small globe; the small monster was standing on the large globe; and the large monster was standing on the medium-sized globe. Because this situation offended their keenly developed sense of symmetry, they proceeded to transfer themselves from one globe to another so that each monster would have a globe proportionate to its own size.

   Monster etiquette complicated the solution of the problem because it requires that:

   1. only one monster may be transferred at a time;
   2. if two monsters are standing on the same globe, only the larger of the two may be transferred;
   3. a monster may not be transferred to a globe on which a larger monster is standing.

   By what sequence of transfers could the monsters have solved this problem?

3. *Monster Problem (Change Form 1)*

   Three five-handed extraterrestrial monsters were holding three crystal globes. Because of the quantum-mechanical peculiarities of their neighborhood, both monsters and globes come in exactly three sizes with no others permitted: small, medium, and large. The medium-sized monster was holding the small globe; the small monster was holding the large globe; and the large monster was holding the medium-sized globe. Because this situation offended their keenly developed sense of symmetry, they proceeded to shrink and expand globes so that each monster would have a globe proportionate to its own size.

   Monster etiquette complicated the solution of the problem because it requires that:

   1. only one globe may be changed at a time;
   2. if two globes have the same size, only the globe held by the larger monster may be changed;
   3. a globe may not be changed to the same size as the globe of a larger monster.

   By what sequence of changes could the monsters have solved this problem?

4. *Monster Problem (Change Form 2)*

   Three five-handed extraterrestrial monsters were holding three crystal globes. Because of the quantum-mechanical peculiarities of their neighborhood, both monsters and globes come in exactly three sizes with no others permitted: small, medium, and large. The medium-sized monster was holding the small globe; the small monster was holding the large globe; and the large monster was holding the medium-sized globe. Because this situation offended their keenly developed sense of symmetry, they proceeded to shrink and expand themselves so that each monster would have a globe proportion to its own size.

Table 23.2
(continued)

Monster etiquette complicated the solution of the problem because it requires that:
1. only one monster may be changed at a time;
2. if two monsters have the same size, only the monster holding the large globe may be changed;
3. a monster may not be changed to the same size as a monster holding a larger globe.
By what sequence of changes could the monsters have solved this problem?

Note: From Hayes, J. R., & Simon, H. A. (1976). Psychological differences among problem isomorphs. In N. Castellan, Jr., D. Pisoni, & G. Potts (Eds.), *Cognitive theory* (Vol. II, pp. 23–24). Potomac, MD: Lawrence Erlbaum Associates.

The possibility that there are several hundred plausible learning and thinking strategies may be an important piece of metacognitive knowledge for teachers and educational researchers. As teachers, this knowledge should lead us to question whether we can expect very much general benefit from teaching any single strategy and to consider instead designing courses that allow students to choose among large numbers of strategies. As educational researchers, the knowledge may lead us to try to simplify the evaluation task by searching for categories of strategies that may be evaluated together.

What can we do to reduce the difficulty that people experience in transferring skills? I offer the following speculation based distantly on observations made by Simon and me: People employ certain fundamental categories when they construct representations. I suggest that the most fundamental ones are object, event, action, location, time, and attribute. When the elements of one problem isomorph fall in the same categories as the corresponding elements of another isomorph, then transfer between the two will be easy. For example, it should be easy to transfer from a problem isomorph in which people are moved among apartments to one in which checkers are moved among board positions, because people and checkers are both objects and apartments and board positions are both locations. However, transfer should be difficult to a third isomorph in which events are shuffled in time, because the categories of the elements of the first two problems are different from those in the third problem.

If this speculation is correct, it would suggest that we should not expect students to transfer knowledge across category boundaries without help. Rather, when full understanding of a principle requires students to generalize across category boundaries, we should be prepared to provide the student with examples that illustrate the application of the principle in each major category.

References

DeGroot, A. D. (1965). *Thought and choice in chess*. The Hague: Mouton.

Grove's dictionary of music and musicians. (1954). J. A. F. Maitland (Ed.). Philadelphia: T. Presser.

Hayes, J. R. (1981). *The complete problem solver*. Philadelphia: The Franklin Institute Press.

Hayes, J. R., & Simon, H. A. (1976). Psychological differences among problem isomorphs. In N. Castellan, Jr., D. Pisoni, & G. Potts (Eds.), *Cognitive theory* (Vol. II). Potomac, MD: Lawrence Erlbaum Associates.

Koechel ABC. (1965). H. Von Hase (Ed.). New York: C. F. Peters Corporation.

Polya, G. (1973). *How to solve it*. (2nd ed.). Princeton, NJ: Princeton University Press.

Schonberg, H. C. (1970). *The lives of the great composers*. New York: Norton.
Schwann-l Record & Tape Guide. (1979, August). Boston: ABC Schwann.
Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist, 61*, 394–403.
Simon, H. A., & Gilmartin, K. (1975). A simulation of memory for chess positions. *Cognitive Psychology, 5*, 29–46.
Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. *Cognitive Psychology, 8*, 165–190.

Chapter 24

Musical Expertise

John A. Sloboda

This chapter treats six connected issues having to do with musical expertise. Section 24.1 examines the difficulties associated with characterizing expertise in a way that offers a genuine foothold for cognitive psychology, and I suggest that expertise may not, in fact, be "special" in any cognitively interesting sense. Section 24.2 goes on to review some experimental studies of music, which suggest that most members of a culture possess tacit musical expertise, expressed in their ability to use high-level structural information in carrying out a variety of perceptual tasks. This expertise seems to be acquired through casual exposure to the musical forms and activities of the culture. Section 24.3 provides two detailed examples of exceptional musical expertise (a musical savant and a jazz musician) that apparently developed in the absence of formal instruction, suggesting that normal and "exceptional" expertise may be parts of a single continuum. The evidence presented in section 24.4 suggests that a major difference between musical expertise and many other forms of expertise is that musical expertise requires an apprehension of a structure–emotion mapping. Without this, the ability to perform with "expression" cannot be acquired. Section 24.5 outlines some evidence to suggest that these structure–emotion links become firmly established during middle childhood, under certain conditions, and that these conditions are predictive of future development of musical expertise. Finally, section 24.6 reviews some research efforts that are attempts to clarify the precise nature of the structure–emotion link and are showing that definite types of structures seem to mediate distinct emotions.

24.1 What Is Expertise?

In beginning to think about how a psychologist who deals with music could contribute in a specific way to a volume on expertise, it became clear to me that most of the recently published work on musical competence has made little attempt to define or characterize musical expertise. What we have, instead, is a varied collection of empirical studies on single aspects of what some musicians do. The topics of such studies range from pitch memory (Ward & Burns, 1982), through synchronization in performance (Rasch, 1988), to planning a composition (Davidson & Welsh, 1988), and it is not immediately clear that such accomplishments have anything in common other than the fact that they

From chapter 6 in *Toward A General Theory of Expertise*, ed. K. A. Ericsson and J. Smith (New York: Cambridge University Press, 1991), 153–171. Reprinted with permission.

are different aspects of handling the organized sounds our various societies label as music.

That observation led back to a logically prior question: Is there anything that all examples of expertise *in general* should or might have in common? More precisely, is there anything about the *internal* psychological structures of certain accomplishments that marks them out as examples of expertise? It is important to remember that when someone is declared an expert, that is a social act that may or may not correspond to an intrinsic characteristic of the person so designated.

One possible definition of an expert is "someone who performs a task significantly better (by some specified criterion) than the majority of people." According to this definition, Chase and Ericsson's (1981) digit memorizer SF is an expert. If, however, digit-span recall became a popular hobby, then he might well be overtaken by sufficient numbers of people so that he would cease to be considered an expert. Such a relativistic attribution of expertise clearly would preclude the possibility of any *cognitive* account of expertise, because the cognitive apparatus that earned SF expert status would remain precisely the same after SF was no longer labeled an expert. It does, however, seem to me that exactly such a relativistic conception underlies much common talk of expertise, and to a certain extent determines the agendas of "expertise" research.

For cognitive psychology to have an authentic foothold, we have to find a characterization of expertise that will allow any number of people (up to and including all) to be expert in a particular area. For instance, many would, I think, agree that the vast majority of people are expert speakers of their native languages. I shall later suggest that the majority of our population possess particular types of musical expertise. A possible definition with this outcome might relate to the reliable attainment of specific goals within a specific domain. So, for instance, one is an expert diner if one can get a wide variety of foodstuffs from plate to mouth without spilling anything.

An apparent problem with this definition, however, is that there is no lower limit to the simplicity or specificity of the task to which one can apply it. For instance, this definition would allow each of us to be expert at pronouncing his or her own name or at folding his or her arms. It may seem that we need more than goal attainment to attribute expertise. For instance, one may want to say that an expert is someone who can make an appropriate response to a situation that contains a degree of unpredictability. So the expert bridge player is one who can work out the play most likely to win with a hand that the player has never seen before; the expert doctor is one who can provide an appropriate diagnosis when faced with a configuration of symptoms never before encountered. In this way we might be able to carve out precisely the set of activities in which various experts have been interested.

On further examination, however, it is not as easy to apply this distinction as it might first appear. Pronouncing one's own name can also be seen as an act requiring the handling of unpredictability. It is an act that is occasioned by cues (external or internal) that can vary. One must be able to retrieve and execute the required motor program regardless of the immediate mental context. The complexity of these apparently simple acts is soon revealed when one attempts

to construct machines that can do the same tasks, as the discipline of artificial intelligence has amply documented (e.g., visual recognition [Marr, 1982]).

It is difficult for me to escape the conclusion that we should abandon the idea that expertise is something special and rare (from a cognitive or biological point of view) and move toward the view that the human organism is in its essence expert. The neonatal brain is already an expert system. "Becoming expert" in socially defined ways is the process of connecting "intrinsic" expertise to the outside world so that it becomes manifest in particular types of behaviors in particular types of situations. I believe that Fodor (1975), from another point of view, was articulating a similar proposal: To broadly paraphrase Fodor, "You can't learn anything you don't already know."

To look at expertise in this way may require reversal of some of our perspectives on familiar situations. For instance, when considering Chase and Ericsson's (1981) study of SF, it is easy to allow one's focus of attention to fall on the two hundred hours of practice that moved him from average to the world's best, implicitly equating the acquisition of the expertise with the work that went on in the practice period under observation. The perspective to which I am increasingly drawn suggests that we focus our attention instead on what SF brought to the experimental situation. SF's intimate knowledge of running times was, from this perspective, the principal manifestation of expertise that "bootstrapped" the digit-span task, and it seems to me that the most interesting psychological considerations are how and why that knowledge came to be applied to the task in hand when it did. What determined that it would be applied after about fifteen hours of practice rather than instantaneously or not at all? A plausible answer to that question may well be "chance" (e.g., a particular sequence of numbers that strongly reminded SF of a well-known running time).

In other words, the broad answer to the question of how SF became expert at the digit-span task is that he was able to increment his expertise by approximately 0.01 percent in a situation in which he was already expert at a number of things, including running times, that supplied the other 99.99 percent of what was needed. And each of those preceding areas of expertise was likewise resting on other forms of expertise in the same relationship in a constant, unbroken sequence back to birth and beyond. What made SF "exceptional" in conventional terms was no more than a unique set of life experiences. In the sections that follow, I pursue some implications of this way of looking at expertise as applied to music.

24.2 Acquiring Musical Skill

One of the principal reasons for studying expertise is practical. Given that it would be socially desirable for certain manifestations of expertise to be more widespread than they are, we want to know what we can do to assist people to acquire them. The issue becomes acute in relation to formal education, where the general perception is that we set up environments that are supposed to encourage expertise, but that many individuals still do not achieve levels that we know to be possible (whether it be learning a foreign language, a musical

instrument, or physics skill). We want to be able to tell teachers that there are principled things that they can do to increase the frequency of those 0.01 percent increments in learning.

Music is no exception to this, and music teachers are continually inquiring of psychologists how psychological insights can inform their work. It is their perception that musical expertise is taught and acquired with great difficulty. They speak of "tone-deaf" children (usually children unable to sing in tune); they speak of the difficulty of teaching sight reading, of teaching rhythm, of teaching good intonation on a string instrument, and so on.

My early research on the skill of sight reading has been summarized elsewhere (Sloboda, 1984). That research was carried out under the influence of the previously published work of Chase and Simon (1973) on chess perception. Their research showed that, like playing chess, reading of music depended on an ability to pick up various sorts of patterns in the stimulus. For instance, good sight readers were found to be much more prone than poor sight readers to a sort of "proofreader's error" (Sloboda, 1976a) whereby notational mistakes out of character with the genre were automatically corrected back to what the genre would have predicted. Their ability to use music structure to "chunk" notes could account for their superior short-term memory for notation (Halpern & Bower, 1982; Sloboda, 1976b).

Encouraging as it was to find results for music that so clearly paralleled Chase's findings, I became progressively more disheartened as I talked about those results to groups of teachers. The question they all asked was of what prescriptions I would draw from my results for the teaching of sight reading, and after some hand waving I really had to admit that there were no prescriptions that I could draw at that time. I did not know how one could teach children to "see" structures.

Since then I have come to realize that in order to "see" musically significant structures, one first must be able to "hear" those structures, and I have learned from reading some excellent recent research that the process of coming to "hear" musical structure is a process that occurs quite naturally for the majority of children as a function of normal enculturation. For instance, Zenatti (1969) showed that children at age 7 show a distinct memory advantage for sequences conforming to rules of normal tonal progression, as compared with atonal sequences. This advantage is not shared by children of age 5. Similar results were obtained from studying children's songs (Dowling, 1982, 1988; Gardner, Davidson, & McKernon, 1981). There is a definite age progression from tonal inconsistency and instability toward conformity to the norms of the tonal culture.

An experiment I conducted earlier (Sloboda, 1985a) showed that the progressing of the ability to discriminate between "legal" and "illegal" sequences did not seem to depend on children's receiving any sort of formal music instruction. Almost no children at age 5 made meaningful discriminations, whereas almost all 11-year-olds made discriminations in accordance with those of adults (and music harmony textbooks). The children who were receiving formal music lessons did not fare better than other children.

Although many experiments with adults have shown cognitive differences between musicians and nonmusicians, some studies have shown little differ-

ence. For instance, Deliege and El Ahmahdi (1990) showed that musicians and nonmusicians were remarkably similar in the segmentations they suggested for an atonal piece. That may have been partly attributable to the relative unfamiliarity of the genre to both groups. More strikingly, Bigand (1990) showed that nonmusicians had an ability similar to that of musicians to classify superficially different conventional tonal melodies into groups containing underlying structural similarities. Studies of memory recall for melodies (Sloboda & Parker, 1985) have shown that musicians and nonmusicians have similar abilities to preserve higher-order structure at the expense of note-to-note detail.

The research literature, therefore, leads to the conclusion that human beings pick up quite high-level implicit (or tacit) knowledge about some major structural features of the music of their culture. They gradually improve their ability to do this over the first ten or more years of life and preserve this ability into adulthood. We may presume that this is achieved through informal engagement in the everyday musical activities that abound in almost all human cultures (e.g., nursery rhymes, hymns, dances, popular songs, playground games). In our own culture these forms are, of course, massively reinforced through the broadcast media.

In this way, almost every member of a culture is a musical expert, but the expertise is usually hidden and tacit. It may not exhibit itself in abilities to sing or play. It is, however, manifested in a variety of perceptual and memory tasks. Nearly all of us can identify some kinds of "wrong notes" when we hear them, even though we cannot always say why the notes are "wrong."

Tacit expertise depends, in part, on being in a culture in which one is exposed to products in the specified domain without the necessity for active engagement. This allows the dissociation between receptive expertise and productive expertise. Such a dissociation would not normally occur in chess, or bridge, or physics, because the only way one normally gets exposure to the relevant structures is by *doing* the activity.

It is not the purpose of this essay to give an account of the various developments in understanding what it is that people know when they know about music structure. Suffice it to say that it seems necessary to postulate mechanisms for representing music that are multidimensional and hierarchical. This means that music can be characterized by points of greater or lesser prominence or distance from one another and that various dimensions may be in synchrony or in opposition. This gives rise to complex patterns of tension and resolution at different hierarchical levels. Some of the most influential characterizations of musical representation have been offered by Lerdahl and Jackendoff (1983), Krumhansl (1990), and Meyer (1973).

More pertinent for our current purposes is the observation that at least some of these structures seem capable of being represented in a connectionist network (Bharucha, 1987). A connectionist model of the brain shows one way in which it might be possible for knowledge of complex structures to be built up simply as a result of frequent exposure to relevant examples. Such an activity seems to be an essential requirement of any mechanism that acquires expertise from environments that are not engineered to be instructional (i.e., most environments).

24.3 Acquisition of Musical Expertise in Noninstructional Settings

Musical expertise, in the foregoing sense, is possessed by the majority of untutored members of any culture. This is not, however, what most people mean when they refer to musical expertise; they mean overt skills of performance or composition. Surely these cannot be acquired other than through formal instruction. It is certain that such skills are acquired mainly through instruction, at least in our culture, but there is some evidence that such instruction is not necessary. Several cases of overt expertise have apparently arisen without any formal tuition or intervention by other experts. An examination of these cases is particularly important if we are to isolate the general conditions for the acquisition of expertise.

24.3.1 Musical Prodigies and Savants

There have been several documented cases of children who showed exceptional precocity at various musical skills. Some of them, such as Mozart, went on to become exceptional adults. Others did not sustain their exceptionality into adult life (see Bamberger, 1986, for a cognitive account of adolescent "burnout" among musical prodigies). One of the fullest accounts of a child musical prodigy was given by Revesz (1925), who made an intensive study of the young Hungarian prodigy Erwin Nyherigazy (EN). Although EN had a great deal of formal tuition and support from professional musicians from an early age, he soon surpassed his teachers in his ability to commit tonal piano music to memory on one or two exposures.

There is another group of prodigies who, by and large, do not receive formal instruction: the so-called idiots savants (see Treffert, 1988, for a review). The savant is a person of generally low IQ, usually male, and often autistic, who has developed a skill in one defined area to a level quite exceptional compared with the general population. Although such cases have been reported in the literature for many years, the reports have mostly been only anecdotal and impressionistic contributions to the psychiatric literature. Only in the past decade have systematic investigations of musical savants been reported in the cognitive literature (e.g., Miller, 1987).

One of these studies concerned the autistic savant NP (Sloboda, Hermelin, & O'Connor, 1985). At the time of detailed study, NP was in his early twenties, and we were able to document his ability to recall a tonal piano movement almost perfectly twelve minutes after first hearing it. Two features of the study were particularly noteworthy: (1) His ability did not extend to a simple atonal piece, and (2) the few errors in his recall of the tonal piece were largely in conformity with the rules of the genre. We concluded that NP's recall ability was predicated on his ability to code and store tonal music in terms of its structural features. In that respect, NP's ability was every bit as "intelligent" as the memory performance of chess masters. Other studies of musical savants (Hermelin, O'Connor, & Lee, 1987; Miller, 1987; Treffert, 1988) have confirmed the importance of structural knowledge in supporting their skills.

Because NP was still relatively young when studied, it was possible to talk to people who knew him at different points in his life and observed his ability develop. It seems that NP's early life was one of considerable cultural depriva-

tion. As far as we know, he had few, if any, opportunities to interact with musical instruments and was not encouraged to sing or to engage with music. His precocity was first noticed at about the age of 6 years, when he spontaneously reproduced at the piano a song that a staff member at his day-care center had just played. From the point on, he was given many opportunities and encouragements to interact with music and musical instruments, although nothing approaching "instruction" was ever possible with this profoundly nonverbal individual. Even now his "lessons" consist of a pianist playing pieces that NP then reproduces. A tape recording of his accomplishments at the age of 8 years shows memory and performance skills that were impressive for an autistic child, though by no means as polished and outstanding as his current performances.

How did NP's skill compare with "normal" skill at the various stages of his life? At age 6 or 7, it was not clear that his memorization abilities were abnormally good. Most untutored children of that age are capable of memorizing short songs, and many can succeed in picking them out on a piano by a process of trial and error. What distinguished NP at that age was his ability to map his internal knowledge of songs directly and without error onto the piano keyboard and to choose appropriate fingering patterns. His performances of tonal music have always been characterized by an absence of hesitation or experimentation, no doubt assisted by his possession of absolute pitch. We have no information that would help us to explain how NP acquired his knowledge without having had any known opportunity to practice before the age of 6.

For the period of his early twenties, the comparison with normals showed a somewhat different pattern. His technical accomplishments were then not unusual. Many reasonably proficient pianists can choose appropriate fingerings for musical passages immediately and automatically. What made NP quite unusual was the *length* of the musical material he could commit to accurate memory after a single hearing. This is a skill shared by few adults at any level of musical expertise, although there are adult musicians of my acquaintance who claim that they could do what NP does when they were age 12 or 13. They no longer can do it, because it has not seemed interesting or worthwhile for them to practice and maintain that particular skill.

We may ask what conditions seem to be associated with the acquisition of the expertise of NP and other savants. The first common factor seems to be a high degree of intrinsic motivation for engagement with a single activity sustained over many years. Such motivation usually has a strong obsessional component, in that given freedom, the savant will spend all available time on the activity, without ever tiring of it.

The second factor is an environment that provides frequent opportunities for the practice of the skill in question. In the case of a musical savant, this may include the provision of regular access to instruments, broadcast media, and musical events. It is possible to suppose that whatever level of cultural deprivation NP suffered during his earliest years, he at least would have been exposed to music through the broadcast media.

The third factor is, of course, the exceptional amount of time spent in cognitive engagement with the materials and activities relevant to the skill in question (practice). It is difficult to estimate the amount of time NP spent thinking

about music when not playing or listening to it, but obvious external involvement probably amounted to four to five hours per day.

The fourth factor, therefore, is the availability of the time and opportunity to "indulge" the obsession. It may be because fewer societal demands are made on people with low IQs that they are "allowed," even encouraged, to devote their attentions in this way.

The fifth factor is the complete absence of negative external reinforcement related to attainment or lack of it. There is, therefore, little possibility of a savant's developing self-doubt, fear of failure, or any of the other blocks that inhibit and sometimes prevent normal or exceptional accomplishment.

24.3.2 Jazz Musicians

It is probable that many of the world's musical cultures, particularly the informal, nonliterate "folk" cultures, have been breeding grounds for expertise. Some anthropological work (e.g., Blacking, 1976) suggests that this is true of indigenous Third World cultures. The jazz culture of New Orleans in the early part of this century may not have been greatly different from those other cultures in many respects. Its advantage for us is that jazz rapidly spread from New Orleans to become part of mass culture and contributed an entirely new facet to the face of Western culture. Its leaders became cult heroes, and jazz itself became a subject for intensive academic scrutiny. For these reasons, we have far more detailed biographical information about jazz musicians than about the musicians from all of the world's other nonliterate cultures put together.

It appears that most of the early jazz players were self-taught. Among the self-taught players who became international names were Bix Beiderbecke, Roy Eldridge, and Louis Armstrong. Collier's (1983) study of Armstrong is particularly detailed, and it allows us to look at Armstrong's musical development in some detail as a "prototype" of untutored expertise.

Armstrong spent most of his early years in a neighborhood known as "Black Storeyville," an area designated for black prostitution. One of the features of that neighborhood was the continual live music, performed by dance bands and "tonk" bands, which often would play on the street to attract customers. Having little knowledge of the world outside, Armstrong had little more than pimps and musicians as male role models. His father had abandoned his mother before he was born. His childhood was one of extreme poverty and deprivation, and from the age of 7 years he had to work, steal, and hustle to make money for his mother and himself. At the age of 8 or 9 years he formed a vocal quartet with some other boys in order to pick up pennies on street corners. The group lasted two or three years and probably practiced and performed in public two or three times per week. That provided several hundred hours of improvised part singing, which as Collier observed, "would have constituted a substantial course in ear training—far more than most conservatory instrumentalists get today."

At the age of 13 or 14 years, Armstrong was involved in an incident with a gun and was, as a result, sent to the Colored Waif's Home (known as the Jones Home). There the boys were taught reading, writing, and arithmetic, with gardening as a sideline. The home had a band that played once a week around the

city. After six months in the home, Armstrong was allowed to join the band, first playing tambourine, then drums, then alto horn. It is clear from contemporary accounts that many of the bands playing in the streets of New Orleans were fairly informal groups with an "anything goes" attitude. It was quite easy for a novice to join in the general noise, just playing the notes he knew, and his mistakes and split notes would pass without comment. Armstrong quickly learned how to get sounds out of the horn, and his vocal experience made it easy for him to work out appropriate parts to the songs the band played. His talent was noticed, and he was promoted to bugle player. He gradually improved to become the band's leader, but he left the home and the band after two years, at age 16. Nothing he experienced in the home would merit the term "formal teaching."

Armstrong found casual work driving a coal cart, which occupied his days, but during the evenings he began playing jazz in the blues bands of the tonks. He did not at that stage own a cornet, and so it was impossible for him to practice. He simply went around to the various bands asking cornetists to let him sit in for a few numbers. Blues music provided a good vehicle for gaining jazz expertise. Blues songs featured slow tempos in two or three of the easiest keys. The set melodies were of the simplest sort; in many cases there was no set melody at all, and the cornetist would string phrases together from a small repertoire of stock figures.

At age 17, Armstrong acquired his first cornet and began to practice and work regularly at one of the tonks. The work paid little, and so he kept his coal job during the day. At some point in that period Armstrong met Joe Oliver, acknowledged as the best cornetist in New Orleans. Armstrong began hanging around the places where Oliver played, running errands, carrying his case, and eventually sitting in for him. Oliver became Armstrong's sponsor and to some extent his teacher. According to Collier, however, Oliver did not influence Armstrong's style and probably did little more than show Armstrong some new tunes and possibly a few alternative fingerings.

By age 19, Armstrong was finding employment on local riverboat excursions. Then, for three summers running, he made long trips, playing every day. For the first time in his life music had become his predominant activity. The band played seven nights per week, doing fourteen numbers and encores each night. They rehearsed two afternoons per week, and the repertoire changed every two weeks. It was only after joining the riverboats that Armstrong learned how to read music and had to acquire the discipline of playing what was written rather than what he felt like playing. When he left the riverboats at age 23, he was an established professional musician.

If Armstrong's early life was a prototype for untutored acquisition of expertise, which of its features might we highlight for future corroboration? One obvious feature was the casual immersion in a rich musical environment with many opportunities to listen and observe. A second feature was the early systematic exploration of a performance medium (in his case, voice). Third, as far as we can judge, his early experiences allowed a great deal of freedom to explore and experiment without negative consequences. A fourth feature was a lack of distinction between "practice" and "performance." The learning took place on the job. A fifth feature was an enduring motivation to engage in

music—in Armstrong's case, a complex mix of internal and external motivations, but arguably with internal motivations dominating. A sixth feature was a graded series of opportunities and challenges available or sought out as the expertise developed.

In many ways, this list of features fits the case of a savant such as NP. The principal differences in the two examples cited here relate to motivation and challenge. NP's motivation did not have a significant external component, and partly for that reason it is not clear that his challenges either arose or were grasped with the same frequency as those of Armstrong. It is easy to imagine NP remaining on a performance plateau. Armstrong went on growing and changing throughout his life.

What these case studies show is that high levels of expertise are achievable without instruction. This does not, of course, mean that instruction is useless. By providing a structured progression of information and challenges for a learner, geared precisely to the learner's capacities at a given time, a teacher may be able to accelerate a learner's progress. Not every person has the opportunity to extract the relevant experiences from the "natural" environment that Armstrong had. A formal instructional environment can engineer the conditions for such extraction. The danger of all such environments is that goals and standards are imposed on the learner, rather than being chosen. The consequence can be to inhibit intrinsic motivation and originality (Amabile, 1983). If external constraints are extreme, it may even be that the ability to enjoy music will be destroyed.

In this connection, one other difference between NP and Armstrong has not been brought out thus far. One of the most striking aspects of NP's musical life was its lack of affect. All pieces in his repertoire were played in a "wooden," unexpressive manner. Although his immediate reproduction showed some of the expressive features of the model, within twenty-four hours all expressive variation was "washed out," leaving a rigid metronomical husk. It was as if NP had no means of understanding (and thus relating to the structure of) the small variations in timing, loudness, and timbre that are the lifeblood of musical performances. From the earliest recording we have of Armstrong's music, in contrast, we find a richly expressive, flexible performance that bends tone and time in ways that have a strong impact on many listeners. Armstrong is not hailed as the king of jazz for his technique, impressive as it was. There are others who match or surpass him in technique. He is revered for the life he could breathe into the simplest material.

NP was one of a rather small number of people who appear to gain complete satisfaction from relating to music as pure structure or syntax. What brings the vast majority of us to music, and keeps us with it, is something additional: its power to mediate a vast range of emotionally toned states, ranging from the subtle to the overwhelming. Because modern systematic studies of music have approached it with the tools of cognitive science and linguistics, the emotional aspect of music has been virtually overlooked, and naive readers of modern research studies might be forgiven for thinking that music is simply another kind of complex structure to be apprehended, like chess or physics.

I know that those who are expert in chess or physics say that there is beauty and emotion in those activities too, but there is a sense in which such things are

not central to the skill. One can write a perfectly effective computer program for chess that will not need any information about how particular chess positions or games will affect the emotions of certain human players. I think there is a strong case for saying that a computer could never adequately simulate Louis Armstrong without some implementation of a theory of the emotions.

24.4 Expression and Emotion as Foundational Aspects of Musical Expertise

Those approaching music with the prejudices and preoccupations of experimental psychology have been wary of examining the emotional aspect, for methodological and conceptual reasons. Rather than examine these reasons in detail, I should like to point to some investigations that seem to have "opened doors" into this area.

The advent of the microcomputer and microtechnology has, for the first time, made possible easy and accurate transfer of detailed performance information into computers for sophisticated analysis. The 1980s saw a number of studies (Clarke, 1985; Gabrielsson, 1983; Shaffer, 1981; Sloboda, 1983; Sundberg, 1988; Todd, 1985) that measured minute expressive variations in performance loudness and timing. These studies showed several things: (1) A given player can consistently repeat given variations on successive performances; (2) these perturbations are not random but, rather, are intentional, and performers can alter them to a greater or lesser extent at will; (3) many of these perturbations are rule-governed and relate to the formal structure of the music in systematic ways.

My own studies (Sloboda, 1983, 1985b), for instance, have shown that timing deformations are organized around the strong metrical beats of tonal melodies in a way that makes the metrical structure clearer for listeners than it is when such deformations are not present. Although we do not yet have the evidence, this line of research suggests that all effective expression may be systematic and rule-governed in this way, helping to highlight musical structures in a way that makes their emotion-bearing content more manifest to listeners.

The other line of contemporary thinking that converges with the experimental work on expression is the music-theory work of such writers as Leonard Meyer (Meyer, 1956, 1973) and Fred Lerdahl (1988a, 1988b; Lerdahl & Jackendoff, 1983). Meyer has convincingly argued that emotion in music arises out of the complex, often subliminal web of expectations and violations of expectations that musical structures unfold over time (Narmour, 1977). Lerdahl (1988b) takes this a step farther by suggesting that only structures that have certain formal properties (such as discreteness and hierarchical organization) can be directly detected by listeners (Balzano, 1980; Shepard, 1982). Only such structures will be effective in creating the types of tensions and resolutions that can support the emotional activities and responses peculiar to music. Lerdahl has particularly enraged certain sections of the avant-garde music community by claiming that traditional tonal music satisfies his criteria, whereas such forms as serial music do not. This could be used as an explanation of why tonality has been able to resist all attempts to oust it from center stage in music and why many avant-garde genres have but limited appeal. The general thrust of all this thinking about music gets independent support from cognitive

theorists (e.g., Ortony, Clore, & Collins, 1988) who characterize the cognitive substrate of all emotion in terms of the violations of various classes of expectations.

These strands of work lead toward the following set of working hypotheses about the vast central bulk of the world's music:

1. One major function of music is to suggest or mediate a range of emotional responses.
2. Common musical structures have particular perceptible properties that support the patterns of expectation underlying such emotions.
3. Expression in musical performance has the effect of making these structural features more prominent, and thus of heightening the emotional response.

24.5 The Roots of Musical Expertise

At the beginning of this chapter, I asked whether all aspects of musical expertise have anything in common. By a rather circuitous route I now come to a proposed answer, which is that they involve apprehension and use of the structure–emotion link. At whatever level, and for whatever activity, what makes the behavior *musically*, as opposed to technically or perceptually, expert is its manifestation of this link. I take it as axiomatic that emotions do not have to be learned (although they may be refined and differentiated through experience). They are part of the "expert system" with which we are born. So what must be learned is how to apprehend those features of musical structures that can be mapped onto and therefore evoke our existing emotions.

Hevner's (1936) pioneering work showed that adult members of a culture generally agree on the emotional characterization of a passage of music, in that they tend to select similar adjectives to describe it (e.g., majestic, gloomy, playful). Gardner (1973) has shown that this ability develops through childhood, with younger children able to use only rather crude descriptions (such as "loud" or "jumpy"). It is, of course, possible that particular kinds of music have come to acquire conventional meanings by routes that do not involve the listener's own emotions. Laboratory studies of people's abilities to *describe* music do not show how these abilities were acquired.

Direct observational studies of children's emotional responses to music have been rare. Moog's (1976) studies showed that preverbal infants could demonstrate quite strong expressions of delight or fear on hearing music. The available evidence suggests that tone quality is the aspect of music that elicits the strongest early reactions. Smooth, treble-register sounds seem to elicit the strongest reactions of attention and pleasure. Most children below the age of 5 years seem not to be particularly interested in unpitched rhythms and seem not to differentiate emotionally between music played in conventional harmony and that played dissonantly.

As children grow older, it is less easy to record emotional responses by direct observation. Socialization leads to significant suppression of direct emotional expression. An alternative approach that I have been pursuing (Sloboda, 1989) is to ask adults to recall musical experiences from the first ten years of life. The

literature on autobiographical memory (Brown & Kulik, 1977; Rubin & Kozin, 1984) suggests that experiences connected with significant emotion may be particularly retrievable. The method also has the advantage of tapping musical experience in a range of naturalistic contexts, rather than in restricted experimental contexts. In addition to asking these adults for information about childhood events and their contexts, I also ask them if those experiences had any particular significance for them. Information about the involvement of music in their lives, including formal music tuition, is also collected.

The findings from these studies indicate that most subjects seem to be capable of producing at least one memory. Some people readily recalled as many as ten different events. No event was recalled from an age earlier than 3 years, but from 4 to 10 years the age spread was fairly even. Analysis of the words used by adults to describe the character of their experiences (both of the music itself and of their reaction to it) showed an interesting age progression. Memories from around age 5 tended to characterize music in rather neutral descriptive terms (e.g., "fast," "loud," "simple"), and the responses to it in terms of general positive enjoyment (e.g., "love," "like," "enjoy," "excited," "happy"). Looking back to age 8, subjects characterized music in terms of its affective or sensual characteristics (e.g., "beautiful," "liquid," "funny"), and the responses to it were recalled in terms of wonder or surprise (e.g., "enthralled," "incredulous," "astounded," "overwhelmed," "awe-struck"). Finally, harking back to around age 9, some memories contained strong feelings of sadness (e.g., "melancholy," "sad," "apprehensive").

It is of particular significance that the ability to respond to music in terms of wonder arises at about the age when children can be shown to distinguish reliably between tonal and atonal music. This strongly suggests that the particular violations of expectations that mediate some of the more "advanced" emotional responses to music require the ability to represent music in terms of the structural categories of tonal music. It is also significant that the progresssion of responsivity seems to owe nothing to explicit formal instruction. The majority of the experiences reported *preceded* the onset of formal musical training, and in several cases such an experience spurred the child to seek instruction. Learning the structure–emotion link seems to proceed in the absence of formal instruction.

Some of the memories reported clearly had the status of what some people call "peak experiences"—unusual and deeply rewarding experiences of a complex emotional/intellectual character. The research showed that people who have had such peak experiences were more likely than others to pursue involvement with music for the rest of their life. The experiences provided a strong source of internal motivation to engage with music in a systematic way (arguably in part to increase the likelihood of replicating the experiences). Educators wishing to raise the general level of musical skill might well be advised to consider how they can help increase the frequency of such experiences in the population, because it is clear that not every child has them.

The memory research provided some interesting clues on this latter point as well. It was discovered that almost none of those peak experiences had occurred in situations of external constraint or anxiety. The most likely environment for a peak experience was at home, on one's own or with friends and

family, and while listening to music. The least promising environment was at school, with teachers, while performing. The individual stories graphically revealed the kinds of anxieties and humiliations many children were made to suffer in relation to music by insensitive adults or through insensitive educational practices. These acted as strong disincentives to further engagement with music and seemed to block the possibility of making links between emotions and the intrinsic characteristics of music.

A similar lesson emerges from a recent study of leading American concert pianists by Sosniak (1989). None of those in her sample showed exceptional promise as a child, but in every case their early lessons were associated with fun and exploration, rather than with practical achievement. It seems that, at least for the crucial early stages of musical development, there is no special strategy we should recommend to educators, other than to stop worrying about particular apparent skill deficiencies and concentrate on not getting in the way of children's enjoyment and exploration of music. In such contexts, children become natural experts who spontaneously seek what they require to bring their expertise to bear on particular practical accomplishments.

24.6 Musical Structure and Emotion

The final question I wish to raise in this chapter concerns the precise nature of the structure–emotion link: What structures elicit what emotions, and why? Although musicologists have long debated this point (e.g., Cooke, 1959; Meyer, 1956), there have been remarkably few attempts to collect empirical data on it. A few physiological studies (e.g., Goldstein, 1980; Nakamura, 1984) have shown that reliable changes in such indices as heart rate and skin conductance can be shown as people listen to specific pieces of music. But such studies generally have not involved subjecting the music itself to detailed structural analysis. A particular characteristic of emotional responses to music is that they often change in nature and intensity over the duration of a piece and are linked to specific events (rather than being a general "wash" of a particular mood). In this respect, they are similar in nature to emotional responses to drama or fiction. To my knowledge, no published studies provide data on the specific points in musical compositions at which intense or peak emotional experiences take place. One problem is that it is difficult to get intersubjective agreement on how to characterize these experiences. Some of my own research entails an attempt to circumvent this problem by asking people to report (retrospectively, at this stage) on the locations in musical compositions at which they reliably experience direct physical manifestations of emotion (e.g., tears, shivers). A significant minority of subjects have been willing and able to do this and have provided a corpus of some 165 "moments" of reliable emotional response. Full details of this study are reported in Sloboda (1991). An analysis of the subset comprising classical instrumental excerpts has revealed three clusters of structural features associated with three different types of responses. These are summarized in table 24.1. This pattern requires confirmation with other types of music and also by direct observation in experimental situations. If confirmed, it will show that many of the emotional responses to music require that the lis-

Table 24.1
Emotion and musical structure

Emotional response	Associated structural features
Tears or lump in throat	Descending circle of 5ths in harmony
	Melodic appoggiatura
	Melodic or harmonic sequence
	Harmonic or melodic acceleration to cadence
Shivers down spine or goose pimples	Enharmonic change
	Delay of final cadence
	New or unprepared harmony
	Sudden dynamic or textural change
Racing heart and "pit of stomach" sensations	Harmonic or melodic acceleration
	Sudden dynamic or textural change
	Repeated syncopation
	Prominent event arriving earlier than expected

tener, at some level, represent high-level structure. For instance, one cannot define "melodic appoggiatura" apart from a description of music in terms of strong and weak beats within a metrical structure and of discord and resolution within a tonal framework. This is one reason we find it difficult to respond emotionally to the music of other cultures as do the members of those cultures. We have not yet assimilated the means of representing their musical structures that would allow the appropriate structure–emotion links to be activated.

We have many interesting and important questions to explore, such as why these particular structures mediate these particular emotions in the way that they do. Research, however, has begun to clarify a major strand in musical expertise that distinguishes it starkly from the other forms of expertise represented in this volume. It suggests that the central conditions for acquisition of musical expertise are as follows:

1. Existence in a musical culture of forms that have perceptible structures of certain kinds (as specified by Lerdahl and others)
2. Frequent informal exposure to examples of these forms over a lifetime
3. Existence of a normal range of human emotional responses
4. Opportunity to experience these emotions mediated through perceived musical structures, which in itself requires
5. Opportunity to experience music in contexts free of externally imposed constraints or negative reinforcements

If we can ensure these conditions, then the problems associated with bringing individuals to levels of achievement we would currently regard as exceptional may turn out to be trivial.

References

Amabile, T. M. (1983). *The social psychology of creativity*. New York: Springer-Verlag.

Balzano, G. J. (1980). The group-theoretic description of twelvefold and microtonal pitch systems. *Computer Music Journal*, *4*, 66–84.

Bamberger, J. (1986). Cognitive issues in the development of musically gifted children. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 388–416). Cambridge: Cambridge University Press.

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception, 5*, 1–30.

Bigand, E. (1990). Abstraction of two forms of underlying structure in a tonal melody. *Psychology of Music, 19*, 45–59.

Blacking, J. (1976). *How musical is man?* London: Faber.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*, 73–99.

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141–189). Hillsdale, NJ: Erlbaum.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 209–236). London: Academic Press.

Collier, J. L. (1983). *Louis Armstrong: An American genius*. New York: Oxford University Press.

Cooke, D. (1959). *The language of music*. London: Oxford University Press.

Davidson, L., & Welsh, P. (1988). From collections to structure: The developmental path of tonal thinking. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation and composition* (pp. 260–285). London: Oxford University Press.

Deliege, I., & El Ahmahdi, A. (1990). Mechanisms of cue extraction in musical groupings: A study of perception, on *Sequenza VI* for viola solo by Luciano Berio. *Psychology of Music, 19*, 18–44.

Dowling, W. J. (1982). Melodic information processing and its development. In D. Deutsch (Ed.), *The psychology of music* (pp. 413–430). New York: Academic Press.

Dowling, W. J. (1988). Tonal structure and children's early learning of music. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation and composition* (pp. 113–128). London: Oxford University Press.

Foder, J. A. (1975). *The language of thought*. Hassocks, Sussex: Harvester Press.

Gabrielsson, A. (1988). Timing in music performance and its relation to music experience. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation and composition* (pp. 27–51). London: Oxford University Press.

Gardner, H. (1973). Children's sensitivity to musical styles. *Merrill-Palmer Quarterly of Behavioral Development, 19*, 67–77.

Gardner, H., Davidson, L., & McKernon, P. (1981). The acquisition of song: A developmental approach. In *Documentary report of the Ann Arbor Symposium*. Music Educators' National Conference, Reston, VA.

Goldstein, A. (1980). Thrills in response to music and other stimuli. *Physiological Psychology, 8*, 126–129.

Halpern, A. R., & Bower, G. H. (1982). Musical expertise and melodic structure in memory for musical notation. *American Journal of Psychology, 95*, 31–50.

Hermelin, B., O'Connor, N., & Lee, S. (1987). Musical inventiveness of five idiots-savants. *Psychological Medicine, 17*, 79–90.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology, 48*, 246–268.

Krumhansl, C. (1990). *Tonal structures and music cognition*. New York: Oxford University Press.

Lerdahl, F. (1988a). Tonal pitch space. *Music Perception, 5*, 315–350.

Lerdahl, F. (1988b). Cognitive constraints on compositional systems. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation and composition* (pp. 231–259). London: Oxford University Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Marr, D. A. (1982). *Vision*. San Francisco: Freeman.

Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Meyer, L. B. (1973). *Explaining music*. Berkeley: University of California Press.

Miller, L. K. (1987). Sensitivity to tonal structure in a developmentally disabled musical savant. *Psychology of Music, 15*, 76–89.

Moog, H. (1976). *The musical experience of the preschool child* (C. Clarke, Trans.). London: Schott.

Nakamura, H. (1984). Effects of musical emotionality upon GSR and respiration rate: The relationship between verbal reports and physiological responses. *Japanese Journal of Psychology*, *55*, 47–50.

Narmour, E. (1977). *Beyond Schenkerism: The need for alternatives in music analysis*. Chicago: University of Chicago Press.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of the emotions*. Cambridge: Cambridge University Press.

Rasch, R. A. (1988). Timing and synchronization in ensemble performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation and composition* (pp. 70–90). London: Oxford University Press.

Revesz, G. (1925). *The psychology of a musical prodigy*. London: Kegan Paul, Trench, & Trubner.

Rubin, D. C., & Kozin, M. (1984). Vivid memories. *Cognition, 16*, 81–95.

Shaffer, L. H. (1981). Performance of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology, 13*, 326–376.

Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *The psychology of music* (pp. 344–390). New York: Academic Press.

Sloboda, J. A. (1976a). The effect of item position on the likelihood of identification by inference in prose reading and music reading. *Canadian Journal of Psychology, 30*, 228–236.

Sloboda, J. A. (1976b). Phrase units as determinants of visual processing in music reading. *British Journal of Psychology, 68*, 117–124.

Sloboda, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology, A35*, 377–396.

Sloboda, J. A. (1984). Experimental studies of music reading: A review. *Music Perception, 2*, 222–236.

Sloboda, J. A. (1985a). *The musical mind: The cognitive psychology of music*. London: Oxford University Press.

Sloboda, J. A. (1985b). Expressive skill in two pianists: Style and effectiveness in music performance. *Canadian Journal of Psychology, 39*, 273–293.

Sloboda, J. A. (1989). Music as a language. In F. Wilson & F. Roehmann (Eds.), *Music and child development: Proceedings of the 1987 Biology of Music Making Conference* (pp. 28–43). St. Louis: MMB Music.

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music, 19* (2), 110–120.

Sloboda, J. A., Hermelin, B., & O'Connor, N. (1985). An exceptional musical memory. *Music Perception, 3*, 155–170.

Sloboda, J. A., & Parker, D. H. H. (1985). Immediate recall of melodies. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 143–167). London: Academic Press.

Sosniak, L. (1989). From tyro to virtuoso: A long-term commitment to learning. In F. Wilson & F. Roehmann (Eds.), *Music and child development: Proceedings of the 1987 Biology of Music Making Conference* (pp. 274–290). St. Louis: MMB Music.

Sundberg, J. (1988). Computer synthesis of musical performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation and composition* (pp. 52–59). London: Oxford University Press.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception, 3*, 33–58.

Treffert, D. A. (1988). The idiot savant: A review of the syndrome. *American Journal of Psychiatry, 145*, 563–572.

Ward, W. D., & Burns, E. M. (1982). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music* (pp. 431–452). New York: Academic Press.

Zenatti, A. (1969). *Le développement génétique de la perception musicale*. Monographies Français Psychologique No. 17.

PART XII

Decision Making

Chapter 25

Judgment under Uncertainty: Heuristics and Biases

Amos Tversky and Daniel Kahneman

Many decisions are based on beliefs concerning the likelihood of uncertain events such as the outcome of an election, the guilt of a defendant, or the future value of the dollar. These beliefs are usually expressed in statements such as "I think that . . . ," "chances are . . . ," "it is unlikely that . . . ," and so forth. Occasionally, beliefs concerning uncertain events are expressed in numerical form as odds or subjective probabilities. What determines such beliefs? How do people assess the probability of an uncertain event or the value of an uncertain quantity? This article shows that people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors.

The subjective assessment of probability resembles the subjective assessment of physical quantities such as distance or size. These judgments are all based on data of limited validity, which are processed according to heuristic rules. For example, the apparent distance of an object is determined in part by its clarity. The more sharply the object is seen, the closer it appears to be. This rule has some validity, because in any given scene the more distant objects are seen less sharply than nearer objects. However, the reliance on this rule leads to systematic errors in the estimation of distance. Specifically, distances are often overestimated when visibility is poor because the contours of objects are blurred. On the other hand, distances are often underestimated when visibility is good because the objects are seen sharply. Thus, the reliance on clarity as an indication of distance leads to common biases. Such biases are also found in the intuitive judgment of probability. This article describes three heuristics that are employed to assess probabilities and to predict values. Biases to which these heuristics lead are enumerated, and the applied and theoretical implications of these observations are discussed.

Representativeness

Many of the probabilistic questions with which people are concerned belong to one of the following types: What is the probability that object A belongs to class B? What is the probability that event A originates from process B? What is the probability that process B will generate event A? In answering such questions, people typically rely on the representativeness heuristic, in which probabilities

From *Science* 185, no. 415 (1974): 1124–1131.

are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B. For example, when A is highly representative of B, the probability that A originates from B is judged to be high. On the other hand, if A is not similar to B, the probability that A originates from B is judged to be low.

For an illustration of judgment by representativeness, consider an individual who has been described by a former neighbor as follows: "Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail." How do people assess the probability that Steve is engaged in a particular occupation from a list of possibilities (for example, farmer, salesman, airline pilot, librarian, or physician)? How do people order these occupations from most to least likely? In the representativeness heuristic, the probability that Steve is a librarian, for example, is assessed by the degree to which he is representative of, or similar to, the stereotype of a librarian. Indeed, research with problems of this type has shown that people order the occupations by probability and by similarity in exactly the same way (Kahneman & Tversky, 1973, 4). This approach to the judgment of probability leads to serious errors, because similarity, or representativeness, is not influenced by several factors that should affect judgments of probability.

Insensitivity to Prior Probability of Outcomes

One of the factors that have no effect on representativeness but should have a major effect on probability is the prior probability, or base-rate frequency, of the outcomes. In the case of Steve, for example, the fact that there are many more farmers than librarians in the population should enter into any reasonable estimate of the probability that Steve is a librarian rather than a farmer. Considerations of base-rate frequency, however, do not affect the similarity of Steve to the stereotypes of librarians and farmers. If people evaluate probability by representativeness, therefore, prior probabilities will be neglected. This hypothesis was tested in an experiment where prior probabilities were manipulated (Kahneman & Tversky, 1973, 4). Subjects were shown brief personality descriptions of several individuals, allegedly sampled at random from a group of 100 professionals—engineers and lawyers. The subjects were asked to assess, for each description, the probability that it belonged to an engineer rather than to a lawyer. In one experimental condition, subjects were told that the group from which the descriptions had been drawn consisted of 70 engineers and 30 lawyers. In another condition, subjects were told that the group consisted of 30 engineers and 70 lawyers. The odds that any particular description belongs to an engineer rather than to a lawyer should be higher in the first condition, where there is a majority of engineers, than in the second condition, where there is a majority of lawyers. Specifically, it can be shown by applying Bayes' rule that the ratio of these odds should be $(.7/.3)^2$, or 5.44, for each description. In a sharp violation of Bayes' rule, the subjects in the two conditions produced essentially the same probability judgments. Apparently, subjects evaluated the likelihood that a particular description belonged to an engineer rather than to a lawyer by the degree to which this description was representative of the two stereotypes, with little or no regard for the prior probabilities of the categories.

The subjects used prior probabilities correctly when they had no other information. In the absence of a personality sketch, they judged the probability that an unknown individual is an engineer to be .7 and .3, respectively, in the two base-rate conditions. However, prior probabilities were effectively ignored when a description was introduced, even when this description was totally uninformative. The responses to the following description illustrate this phenomenon:

> Dick is a 30 year old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues.

This description was intended to convey no information relevant to the question of whether Dick is an engineer or a lawyer. Consequently, the probability that Dick is an engineer should equal the proportion of engineers in the group, as if no description had been given. The subjects, however, judged the probability of Dick being an engineer to be .5 regardless of whether the stated proportion of engineers in the group was .7 or .3. Evidently, people respond differently when given no evidence and when given worthless evidence. When no specific evidence is given, prior probabilities are properly utilized; when worthless evidence is given, prior probabilities are ignored (Kahneman & Tversky, 1973, 4).

Insensitivity to Sample Size

To evaluate the probability of obtaining a particular result in a sample drawn from a specified population, people typically apply the representativeness heuristic. That is, they assess the likelihood of a sample result, for example, that the average height in a random sample of ten men will be 6 feet (180 centimeters), by the similarity of this result to the corresponding parameter (that is, to the average height in the population of men). The similarity of a sample statistic to a population parameter does not depend on the size of the sample. Consequently, if probabilities are assessed by representativeness, then the judged probability of a sample statistic will be essentially independent of sample size. Indeed, when subjects assessed the distributions of average height for samples of various sizes, they produced identical distributions. For example, the probability of obtaining an average height greater than 6 feet was assigned the same value for samples of 1000, 100, and 10 men (Kahneman & Tversky, 1972, 3). Moreover, subjects failed to appreciate the role of sample size even when it was emphasized in the formulation of the problem. Consider the following question:

> A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50 percent, sometimes lower.
>
> For a period of 1 year, each hospital recorded the days on which more than 60 percent of the babies born were boys. Which hospital do you think recorded more such days?

The larger hospital (21)
The smaller hospital (21)
About the same (that is, within 5 percent of each other) (53)

The values in parentheses are the number of undergraduate students who chose each answer.

Most subjects judged the probability of obtaining more than 60 percent boys to be the same in the small and in the large hospital, presumably because these events are described by the same statistic and are therefore equally representative of the general population. In contrast, sampling theory entails that the expected number of days on which more than 60 percent of the babies are boys is much greater in the small hospital than in the large one, because a large sample is less likely to stray from 50 percent. This fundamental notion of statistics is evidently not part of people's repertoire of intuitions.

A similar insensitivity to sample size has been reported in judgments of posterior probability, that is, of the probability that a sample has been drawn from one population rather than from another. Consider the following example:

> Imagine an urn filled with balls, of which $\frac{2}{3}$ are of one color and $\frac{1}{3}$ of another. One individual has drawn 5 balls from the urn, and found that 4 were red and 1 was white. Another individual has drawn 20 balls and found that 12 were red and 8 were white. Which of the two individuals should feel more confident that the urn contains $\frac{2}{3}$ red balls and $\frac{1}{3}$ white balls, rather than the opposite? What odds should each individual give?

In this problem, the correct posterior odds are 8 to 1 for the 4:1 sample and 16 to 1 for the 12:8 sample, assuming equal prior probabilities. However, most people feel that the first sample provides much stronger evidence for the hypothesis that the urn is predominantly red, because the proportion of red balls is larger in the first than in the second sample. Here again, intuitive judgments are dominated by the sample proportion and are essentially unaffected by the size of the sample, which plays a crucial role in the determination of the actual posterior odds (Kahneman & Tversky, 1972). In addition, intuitive estimates of posterior odds are far less extreme than the correct values. The underestimation of the impact of evidence has been observed repeatedly in problems of this type (W. Edwards, 1968, 25; Slovic & Lichtenstein, 1971). It has been labeled "conservatism."

Misconceptions of Chance

People expect that a sequence of events generated by a random process will represent the essential characteristics of that process even when the sequence is short. In considering tosses of a coin for heads or tails, for example, people regard the sequence H-T-H-T-T-H to be more likely than the sequence H-H-H-T-T-T, which does not appear random, and also more likely than the sequence H-H-H-H-T-H, which does not represent the fairness of the coin (Kahneman & Tversky, 1972b, 3). Thus, people expect that the essential characteristics of the process will be represented, not only globally in the entire sequence, but also locally in each of its parts. A locally representative sequence, however, deviates systematically from chance expectation: it contains too many alternations and

too few runs. Another consequence of the belief in local representativeness is the well-known gambler's fallacy. After observing a long run of red on the roulette wheel, for example, most people erroneously believe that black is now due, presumably because the occurence of black will result in a more representative sequence than the occurrence of an additional red. Chance is commonly viewed as a self-correcting process in which a deviation in one direction induces a deviation in the opposite direction to restore the equilibrium. In fact, deviations are not "corrected" as a chance process unfolds, they are merely diluted.

Misconceptions of chance are not limited to naive subjects. A study of the statistical intuitions of experienced research psychologists (Tversky & Kahneman, 1971, 2) revealed a lingering belief in what may be called the "law of small numbers," according to which even small samples are highly representative of the populations from which they are drawn. The responses of these investigators reflected the expectation that a valid hypothesis about a population will be represented by a statistically significant result in a sample—with little regard for its size. As a consequence, the researchers put too much faith in the results of small samples and grossly overestimated the replicability of such results. In the actual conduct of research, this bias leads to the selection of samples of inadequate size and to overinterpretation of findings.

Insensitivity to Predictability

People are sometimes called upon to make such numerical predictions as the future value of a stock, the demand for a commodity, or the outcome of a football game. Such predictions are often made by representativeness. For example, suppose one is given a description of a company and is asked to predict its future profit. If the description of the company is very favorable, a very high profit will appear most representative of that description; if the description is mediocre, a mediocre performance will appear most representative. The degree to which the description is favorable is unaffected by the reliability of that description or by the degree to which it permits accurate prediction. Hence, if people predict solely in terms of the favorableness of the description, their predictions will be insensitive to the reliability of the evidence and to the expected accuracy of the prediction.

This mode of judgment violates the normative statistical theory in which the extremeness and the range of predictions are controlled by considerations of predictability. When predictability is nil, the same prediction should be made in all cases. For example, if the descriptions of companies provide no information relevant to profit, then the same value (such as average profit) should be predicted for all companies. If predictability is perfect, of course, the values predicted will match the actual values and the range of predictions will equal the range of outcomes. In general, the higher the predictability, the wider the range of predicted values.

Several studies of numerical prediction have demonstrated that intuitive predictions violate this rule, and that subjects show little or no regard for considerations of predictability (Kahneman & Tversky, 1973, 4). In one of these studies, subjects were presented with several paragraphs, each describing the performance of a student teacher during a particular practice lesson. Some

subjects were asked to *evaluate* the quality of the lesson described in the paragraph in percentile scores, relative to a specified population. Other subjects were asked to *predict*, also in percentile scores, the standing of each student teacher 5 years after the practice lesson. The judgments made under the two conditions were identical. That is, the prediction of a remote criterion (success of a teacher after 5 years) was identical to the evaluation of the information on which the prediction was based (the quality of the practice lesson). The students who made these predictions were undoubtedly aware of the limited predictability of teaching competence on the basis of a single trial lesson 5 years earlier; nevertheless, their predictions were as extreme as their evaluations.

The Illusion of Validity

As we have seen, people often predict by selecting the outcome (for example, an occupation) that is most representative of the input (for example, the description of a person). The confidence they have in their prediction depends primarily on the degree of representativeness (that is, on the quality of the match between the selected outcome and the input) with little or no regard for the factors that limit predictive accuracy. Thus, people express great confidence in the prediction that a person is a librarian when given a description of his personality which matches the stereotype of librarians, even if the description is scanty, unreliable, or outdated. The unwarranted confidence which is produced by a good fit between the predicted outcome and the input information may be called the illusion of validity. This illusion persists even when the judge is aware of the factors that limit the accuracy of his predictions. It is a common observation that psychologists who conduct selection interviews often experience considerable confidence in their predictions, even when they know of the vast literature that shows selection interviews to be highly fallible. The continued reliance on the clinical interview for selection, despite repeated demonstrations of its inadequacy, amply attests to the strength of this effect.

The internal consistency of a pattern of inputs is a major determinant of one's confidence in predictions based on these inputs. For example, people express more confidence in predicting the final grade-point average of a student whose first-year record consists entirely of B's than in predicting the grade-point average of a student whose first-year record includes many A's and C's. Highly consistent patterns are most often observed when the input variables are highly redundant or correlated. Hence, people tend to have great confidence in predictions based on redundant input variables. However, an elementary result in the statistics of correlation asserts that, given input variables of stated validity, a prediction based on several such inputs can achieve higher accuracy when they are independent of each other than when they are redundant or correlated. Thus, redundancy among inputs decreases accuracy even as it increases confidence, and people are often confident in predictions that are quite likely to be off the mark (Kahneman & Tversky, 1973, 4).

Misconceptions of Regression

Suppose a large group of children has been examined on two equivalent versions of an aptitude test. If one selects ten children from among those who did best on one of the two versions, he will usually find their performance on the

second version to be somewhat disappointing. Conversely, if one selects ten children from among those who did worst on one version, they will be found, on the average, to do somewhat better on the other version. More generally, consider two variables X and Y which have the same distribution. If one selects individuals whose average X score deviates from the mean of X by k units, then the average of their Y scores will usually deviate from the mean of Y by less than k units. These observations illustrate a general phenomenon known as regression toward the mean, which was first documented by Galton more than 100 years ago.

In the normal course of life, one encounters many instances of regression toward the mean, in the comparison of the height of fathers and sons, of the intelligence of husbands and wives, or of the performance of individuals on consecutive examinations. Nevertheless, people do not develop correct intuitions about this phenomenon. First, they do not expect regression in many contexts where it is bound to occur. Second, when they recognize the occurrence of regression, they often invent spurious causal explanations for it (Kahneman & Tversky, 1973, 4). We suggest that the phenomenon of regression remains elusive because it is incompatible with the belief that the predicted outcome should be maximally representative of the input, and, hence, that the value of the outcome variable should be as extreme as the value of the input variable.

The failure to recognize the import of regression can have pernicious consequences, as illustrated by the following observation (Kahneman & Tversky, 1973, 4). In a discussion of flight training, experienced instructors noted that praise for an exceptionally smooth landing is typically followed by a poorer landing on the next try, while harsh criticism after a rough landing is usually followed by an improvement on the next try. The instructors concluded that verbal rewards are detrimental to learning, while verbal punishments are beneficial, contrary to accepted psychological doctrine. This conclusion is unwarranted because of the presence of regression toward the mean. As in other cases of repeated examination, an improvement will usually follow a poor performance and a deterioration will usually follow an outstanding performance, even if the instructor does not respond to the trainee's achievement on the first attempt. Because the instructors had praised their trainees after good landings and admonished them after poor ones, they reached the erroneous and potentially harmful conclusion that punishment is more effective than reward.

Thus, the failure to understand the effect of regression leads one to overestimate the effectiveness of punishment and to underestimate the effectiveness of reward. In social interaction, as well as in training, rewards are typically administered when performance is good, and punishments are typically administered when performance is poor. By regression alone, therefore, behavior is most likely to improve after punishment and most likely to deteriorate after reward. Consequently, the human condition is such that, by chance alone, one is most often rewarded for punishing others and most often punished for rewarding them. People are generally not aware of this contingency. In fact, the elusive role of regression in determining the apparent consequences of reward and punishment seems to have escaped the notice of students of this area.

Availability

There are situations in which people assess the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind. For example, one may assess the risk of heart attack among middle-aged people by recalling such occurrences among one's acquaintances. Similarly, one may evaluate the probability that a given business venture will fail by imagining various difficulties it could encounter. This judgmental heuristic is called availability. Availability is a useful clue for assessing frequency or probability, because instances of large classes are usually reached better and faster than instances of less frequent classes. However, availability is affected by factors other than frequency and probability. Consequently, the reliance on availability leads to predictable biases, some of which are illustrated below.

Biases Due to the Retrievability of Instances

When the size of a class is judged by the availability of its instances, a class whose instances are easily retrieved will appear more numerous than a class of equal frequency whose instances are less retrievable. In an elementary demonstration of this effect, subjects heard a list of well-known personalities of both sexes and were subsequently asked to judge whether the list contained more names of men than of women. Different lists were presented to different groups of subjects. In some of the lists the men were relatively more famous than the women, and in others the women were relatively more famous than the men. In each of the lists, the subjects erroneously judged that the class (sex) that had the more famous personalities was the more numerous (Tversky & Kahneman, 1973, 11).

In addition to familiarity, there are other factors, such as salience, which affect the retrievability of instances. For example, the impact of seeing a house burning on the subjective probability of such accidents is probably greater than the impact of reading about a fire in the local paper. Furthermore, recent occurrences are likely to be relatively more available than earlier occurrences. It is a common experience that the subjective probability of traffic accidents rises temporarily when one sees a car overturned by the side of the road.

Biases Due to the Effectiveness of a Search Set

Suppose one samples a word (of three letters or more) at random from an English text. Is it more likely that the word starts with *r* or that *r* is the third letter? People approach this problem by recalling words that begin with *r* (road) and words that have *r* in the third position (car) and assess the relative frequency by the ease with which words of the two types come to mind. Because it is much easier to search for words by their first letter than by their third letter, most people judge words that begin with a given consonant to be more numerous than words in which the same consonant appears in the third position. They do so even for consonants, such as *r* or *k*, that are more frequent in the third position than in the first (Tversky & Kahneman, 1973, 11).

Different tasks elicit different search sets. For example, suppose you are asked to rate the frequency with which abstract words (*thought*, *love*) and concrete words (*door*, *water*) appear in written English. A natural way to answer

this question is to search for contexts in which the word could appear. It seems easier to think of contexts in which an abstract concept is mentioned (*love* in love stories) than to think of contexts in which a concrete word (such as *door*) is mentioned. If the frequency of words is judged by the availability of the contexts in which they appear, abstract words will be judged as relatively more numerous than concrete words. This bias has been observed in a study (Galbraith & Underwood, 1973) which showed that the judged frequency of occurrence of abstract words was much higher than that of concrete words, equated in objective frequency. Abstract words were also judged to appear in a much greater variety of contexts than concrete words.

Biases of Imaginability

Sometimes one has to assess the frequency of a class whose instances are not stored in memory but can be generated according to a given rule. In such situations, one typically generates several instances and evaluates frequency or probability by the ease with which the relevant instances can be constructed. However, the ease of constructing instances does not always reflect their actual frequency, and this mode of evaluation is prone to biases. To illustrate, consider a group of 10 people who form committees of k members, $2 \leq k \leq 8$. How many different committees of k members can be formed? The correct answer to this problem is given by the binomial coefficient $\binom{10}{k}$ which reaches a maximum of 252 for $k = 5$. Clearly, the number of committees of k members equals the number of committees of $(10 - k)$ members, because any committee of k members defines a unique group of $(10 - k)$ nonmembers.

One way to answer this question without computation is to mentally construct committees of k members and to evaluate their number by the ease with which they come to mind. Committees of few members, say 2, are more available than committees of many members, say 8. The simplest scheme for the construction of committees is a partition of the group into disjoint sets. One readily sees that it is easy to construct five disjoint committees of 2 members, while it is impossible to generate even two disjoint committees of 8 members. Consequently, if frequency is assessed by imaginability, or by availability for construction, the small committees will appear more numerous than larger committees, in contrast to the correct bell-shaped function. Indeed, when naive subjects were asked to estimate the number of distinct committees of various sizes, their estimates were a decreasing monotonic function of committee size (Tversky & Kahneman, 1973, 11). For example, the median estimate of the number of committees of 2 members was 70, while the estimate for committees of 8 members was 20 (the correct answer is 45 in both cases).

Imaginability plays an important role in the evaluation of probabilities in real-life situations. The risk involved in an adventurous expedition, for example, is evaluated by imagining contingencies with which the expedition is not equipped to cope. If many such difficulties are vividly portrayed, the expedition can be made to appear exceedingly dangerous, although the ease with which disasters are imagined need not reflect their actual likelihood. Conversely, the risk involved in an undertaking may be grossly underestimated if some possible dangers are either difficult to conceive of, or simply do not come to mind.

Illusory Correlation

Chapman and Chapman (1969) have described an interesting bias in the judgment of the frequency with which two events co-occur. They presented naive judges with information concerning several hypothetical mental patients. The data for each patient consisted of a clinical diagnosis and a drawing of a person made by the patient. Later the judges estimated the frequency with which each diagnosis (such as paranoia or suspiciousness) had been accompanied by various features of the drawing (such as peculiar eyes). The subjects markedly overestimated the frequency of co-occurrence of natural associates, such as suspiciousness and peculiar eyes. This effect was labeled illusory correlation. In their erroneous judgments of the data to which they had been exposed, naive subjects "rediscovered" much of the common, but unfounded, clinical lore concerning the interpretation of the draw-a-person test. The illusory correlation effect was extremely resistant to contradictory data. It persisted even when the correlation between symptom and diagnosis was actually negative, and it prevented the judges from detecting relationships that were in fact present.

Availability provides a natural account for the illusory-correlation effect. The judgment of how frequently two events co-occur could be based on the strength of the associative bond between them. When the association is strong, one is likely to conclude that the events have been frequently paired. Consequently, strong associates will be judged to have occurred together frequently. According to this view, the illusory correlation between suspiciousness and peculiar drawing of the eyes, for example, is due to the fact that suspiciousness is more readily associated with the eyes than with any other part of the body.

Lifelong experience has taught us that, in general, instances of large classes are recalled better and faster than instances of less frequent classes; that likely occurrences are easier to imagine than unlikely ones; and that the associative connections between events are strengthened when the events frequently co-occur. As a result, man has at his disposal a procedure (the availability heuristic) for estimating the numerosity of a class, the likelihood of an event, or the frequency of co-occurrences, by the ease with which the relevant mental operations of retrieval, construction, or association can be performed. However, as the preceding examples have demonstrated, this valuable estimation procedure results in systematic errors.

Adjustment and Anchoring

In many situations, people make estimates by starting from an initial value that is adjusted to yield the final answer. The initial value, or starting point, may be suggested by the formulation of the problem, or it may be the result of a partial computation. In either case, adjustments are typically insufficient (Slovic & Lichtenstein, 1971). That is, different starting points yield different estimates, which are biased toward the initial values. We call this phenomenon anchoring.

Insufficient Adjustment

In a demonstration of the anchoring effect, subjects were asked to estimate various quantities, stated in percentages (for example, the percentage of African countries in the United Nations). For each quantity, a number between 0

and 100 was determined by spinning a wheel of fortune in the subjects' presence. The subjects were instructed to indicate first whether that number was higher or lower than the value of the quantity, and then to estimate the value of the quantity by moving upward or downward from the given number. Different groups were given different numbers for each quantity, and these arbitrary numbers had a marked effect on estimates. For example, the median estimates of the percentage of African countries in the United Nations were 25 and 45 for groups that received 10 and 65, respectively, as starting points. Payoffs for accuracy did not reduce the anchoring effect.

Anchoring occurs not only when the starting point is given to the subject, but also when the subject bases his estimate on the result of some incomplete computation. A study of intuitive numerical estimation illustrates this effect. Two groups of high school students estimated, within 5 seconds, a numerical expression that was written on the blackboard. One group estimated the product

$$8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$$

while another group estimated the product

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$$

To rapidly answer such questions, people may perform a few steps of computation and estimate the product by extrapolation or adjustment. Because adjustments are typically insufficient, this procedure should lead to underestimation. Furthermore, because the result of the first few steps of multiplication (performed from left to right) is higher in the descending sequence than in the ascending sequence, the former expression should be judged larger than the latter. Both predictions were confirmed. The median estimate for the ascending sequence was 512, while the median estimate for the descending sequence was 2,250. The correct answer is 40,320.

Biases in the Evaluation of Conjunctive and Disjunctive Events

In a recent study by Bar-Hillel (1973) subjects were given the opportunity to bet on one of two events. Three types of events were used: (i) simple events, such as drawing a red marble from a bag containing 50 percent red marbles and 50 percent white marbles; (ii) conjunctive events, such as drawing a red marble seven times in succession, with replacement, from a bag containing 90 percent red marbles and 10 percent white marbles; and (iii) disjunctive events, such as drawing a red marble at least once in seven successive tries, with replacement, from a bag containing 10 percent red marbles and 90 percent white marbles. In this problem, a significant majority of subjects preferred to bet on the conjunctive event (the probability of which is .48) rather than on the simple event (the probability of which is .50). Subjects also preferred to bet on the simple event rather than on the disjunctive event, which has a probability of .52. Thus, most subjects bet on the less likely event in both comparisons. This pattern of choices illustrates a general finding. Studies of choice among gambles and of judgments of probability indicate that people tend to overestimate the probability of conjunctive events (Cohen, Chesnick, & Haran, 1972, 24) and to underestimate the probability of disjunctive events. These biases are readily explained as effects of anchoring. The stated probability of the elementary event (success at

any one stage) provides a natural starting point for the estimation of the probabilities of both conjunctive and disjunctive events. Since adjustment from the starting point is typically insufficient, the final estimates remain too close to the probabilities of the elementary events in both cases. Note that the overall probability of a conjunctive event is lower than the probability of each elementary event, whereas the overall probability of a disjunctive event is higher than the probability of each elementary event. As a consequence of anchoring, the overall probability will be overestimated in conjunctive problems and underestimated in disjunctive problems.

Biases in the evaluation of compound events are particularly significant in the context of planning. The successful completion of an undertaking, such as the development of a new product, typically has a conjunctive character: for the undertaking to succeed, each of a series of events must occur. Even when each of these events is very likely, the overall probability of success can be quite low if the number of events is large. The general tendency to overestimate the probability of conjunctive events leads to unwarranted optimism in the evaluation of the likelihood that a plan will succeed or that a project will be completed on time. Conversely, disjunctive structures are typically encountered in the evaluation of risks. A complex system, such as a nuclear reactor or a human body, will malfunction if any of its essential components fails. Even when the likelihood of failure in each component is slight, the probability of an overall failure can be high if many components are involved. Because of anchoring, people will tend to underestimate the probabilities of failure in complex systems. Thus, the direction of the anchoring bias can sometimes be inferred from the structure of the event. The chain-like structure of conjunctions leads to overestimation, the funnel-like structure of disjunctions leads to underestimation.

Anchoring in the Assessment of Subjective Probability Distributions

In decision analysis, experts are often required to express their beliefs about a quantity, such as the value of the Dow-Jones average on a particular day, in the form of a probability distribution. Such a distribution is usually constructed by asking the person to select values of the quantity that correspond to specified percentiles of his subjective probability distribution. For example, the judge may be asked to select a number, X_{90}, such that his subjective probability that this number will be higher than the value of the Dow-Jones average is .90. That is, he should select the value X_{90} so that he is just willing to accept 9 to 1 odds that the Dow-Jones average will not exceed it. A subjective probability distribution for the value of the Dow-Jones average can be constructed from several such judgments corresponding to different percentiles.

By collecting subjective probability distributions for many different quantities, it is possible to test the judge for proper calibration. A judge is properly (or externally) calibrated in a set of problems if exactly Π percent of the true values of the assessed quantities fall below his stated values of X_{Π}. For example, the true values should fall below X_{01} for 1 percent of the quantities and above X_{99} for 1 percent of the quantities. Thus, the true values should fall in the confidence interval between X_{01} and X_{99} on 98 percent of the problems.

Several investigators (Alpert & Raiffa, 1969, 21; Staël von Holstein, 1971; Winkler, 1967) have obtained probability disruptions for many quantities from

a large number of judges. These distributions indicated large and systematic departures from proper calibration. In most studies, the actual values of the assessed quantities are either smaller than X_{01} or greater than X_{99} for about 30 percent of the problems. That is, the subjects state overly narrow confidence intervals which reflect more certainty than is justified by their knowledge about the assessed quantities. This bias is common to naive and to sophisticated subjects, and it is not eliminated by introducing proper scoring rules, which provide incentives for external calibration. This effect is attributable, in part at least, to anchoring.

To select X_{90} for the value of the Dow-Jones average, for example, it is natural to begin by thinking about one's best estimate of the Dow-Jones and to adjust this value upward. If this adjustment—like most others—is insufficient, then X_{90} will not be sufficiently extreme. A similar anchoring effect will occur in the selection of X_{10}, which is presumably obtained by adjusting one's best estimate downward. Consequently, the confidence interval between X_{10} and X_{90} will be too narrow, and the assessed probability distribution will be too tight. In support of this interpretation it can be shown that subjective probabilities are systematically altered by a procedure in which one's best estimate does not serve as an anchor.

Subjective probability distributions for a given quantity (the Dow-Jones average) can be obtained in two different ways: (i) by asking the subject to select values of the Dow-Jones that correspond to specified percentiles of his probability distribution and (ii) by asking the subject to assess the probabilities that the true value of the Dow-Jones will exceed some specified values. The two procedures are formally equivalent and should yield identical distributions. However, they suggest different modes of adjustment from different anchors. In procedure (i), the natural starting point is one's best estimate of the quality. In procedure (ii), on the other hand, the subject may be anchored on the value stated in the question. Alternatively, he may be anchored on even odds, or 50–50 chances, which is a natural starting point in the estimation of likelihood. In either case, procedure (ii) should yield less extreme odds than procedure (i).

To contrast the two procedures, a set of 24 quantities (such as the air distance from New Delhi to Peking) was presented to a group of subjects who assessed either X_{10} or X_{90} for each problem. Another group of subjects received the median judgment of the first group for each of the 24 quantities. They were asked to assess the odds that each of the given values exceeded the true value of the relevant quantity. In the absence of any bias, the second group should retrieve the odds specified to the first group, that is, 9:1. However, if even odds or the stated value serve as anchors, the odds of the second group should be less extreme, that is, closer to 1:1. Indeed, the median odds stated by this group, across all problems, were 3:1. When the judgments of the two groups were tested for external calibration, it was found that subjects in the first group were too extreme, in accord with earlier studies. The events that they defined as having a probability of .10 actually obtained in 24 percent of the cases. In contrast, subjects in the second group were too conservative. Events to which they assigned an average probability of .34 actually obtained in 26 percent of the cases. These results illustrate the manner in which the degree of calibration depends on the procedure of elicitation.

Discussion

This article has been concerned with cognitive biases that stem from the reliance on judgmental heuristics. These biases are not attributable to motivational effects such as wishful thinking or the distortion of judgments by payoffs and penalties. Indeed, several of the severe errors of judgment reported earlier occurred despite the fact that subjects were encouraged to be accurate and were rewarded for the correct answers (Kahneman & Tversky, 1972, 3; Tversky & Kahneman, 1973, 11).

The reliance on heuristics and the prevalence of biases are not restricted to laymen. Experienced researchers are also prone to the same biases—when they think intuitively. For example, the tendency to predict the outcome that best represents the data, with insufficient regard for prior probability, has been observed in the intuitive judgments of individuals who have had extensive training in statistics (Kahneman & Tversky, 1973, 4; Tversky & Kahneman, 1971, 2). Although the statistically sophisticated avoid elementary errors, such as the gambler's fallacy, their intuitive judgments are liable to similar fallacies in more intricate and less transparent problems.

It is not surprising that useful heuristics such as representativeness and availability are retained, even though they occasionally lead to errors in prediction or estimation. What is perhaps surprising is the failure of people to infer from lifelong experience such fundamental statistical rules as regression toward the mean, or the effect of sample size on sampling variability. Although everyone is exposed, in the normal course of life, to numerous examples from which these rules could have been induced, very few people discover the principles of sampling and regression on their own. Statistical principles are not learned from everyday experience because the relevant instances are not coded appropriately. For example, people do not discover that successive lines in a text differ more in average word length than do successive pages, because they simply do not attend to the average word length of individual lines or pages. Thus, people do not learn the relation between sample size and sampling variability, although the data for such learning are abundant.

The lack of an appropriate code also explains why people usually do not detect the biases in their judgments of probability. A person could conceivably learn whether his judgments are externally calibrated by keeping a tally of the proportion of events that actually occur among those to which he assigns the same probability. However, it is not natural to group events by their judged probability. In the absence of such grouping it is impossible for an individual to discover, for example, that only 50 percent of the predictions to which he has assigned a probability of .9 or higher actually come true.

The empirical analysis of cognitive biases has implications for the theoretical and applied role of judged probabilities. Modern decision theory (de Finetti, 1968; Savage, 1954) regards subjective probability as the quantified opinion of an idealized person. Specifically, the subjective probability of a given event is defined by the set of bets about this event that such a person is willing to accept. An internally consistent, or coherent, subjective probability measure can be derived for an individual if his choices among bets satisfy certain principles, that is, the axioms of the theory. The derived probability is subjective in the

sense that different individuals are allowed to have different probabilities for the same event. The major contribution of this approach is that it provides a rigorous subjective interpretation of probability that is applicable to unique events and is embedded in a general theory of rational decision.

It should perhaps be noted that, while subjective probabilities can sometimes be inferred from preferences among bets, they are normally not formed in this fashion. A person bets on team A rather than on team B because he believes that team A is more likely to win; he does not infer this belief from his betting preferences. Thus, in reality, subjective probabilities determine preferences among bets and are not derived from them, as in the axiomatic theory of rational decision (Savage, 1954).

The inherently subjective nature of probability has led many students to the belief that coherence, or internal consistency, is the only valid criterion by which judged probabilities should be evaluated. From the standpoint of the formal theory of subjective probability, any set of internally consistent probability judgments is as good as any other. This criterion is not entirely satisfactory, because an internally consistent set of subjective probabilities can be incompatible with other beliefs held by the individual. Consider a person whose subjective probabilities for all possible outcomes of a coin-tossing game reflect the gambler's fallacy. That is, his estimate of the probability of tails on a particular toss increases with the number of consecutive heads that preceded that toss. The judgments of such a person could be internally consistent and therefore acceptable as adequate subjective probabilities according to the criterion of the formal theory. These probabilities, however, are incompatible with the generally held belief that a coin has no memory and is therefore incapable of generating sequential dependencies. For judged probabilities to be considered adequate, or rational, internal consistency is not enough. The judgments must be compatible with the entire web of beliefs held by the individual. Unfortunately, there can be no simple formal procedure for assessing the compatibility of a set of probability judgments with the judge's total system of beliefs. The rational judge will nevertheless strive for compatibility, even though internal consistency is more easily achieved and assessed. In particular, he will attempt to make his probability judgments compatible with his knowledge about the subject matter, the laws of probability, and his own judgmental heuristics and biases.

Summary

This chapter described three heuristics that are employed in making judgments under uncertainty: (i) representativeness, which is usually employed when people are asked to judge the probability that an object or event A belongs to class or process B; (ii) availability of instances or scenarios, which is often employed when people are asked to assess the frequency of a class or the plausibility of a particular development; and (iii) adjustment from an anchor, which is usually employed in numerical prediction when a relevant value is available. These heuristics are highly economical and usually effective, but they lead to systematic and predictable errors. A better understanding of these

heuristics and of the biases to which they lead could improve judgments and decisions in situations of uncertainty.

Acknowledgments

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Office of Naval Research under contract N00014-73-C-0438 to the Oregon Research Institute, Eugene. Additional Support for this research was provided by the Research and Development Authority of the Hebrew University, Jerusalem, Israel.

References

Alpert, M., and H. Raiffa. "Unpublished manuscript."

Bar-Hillel, M. (1973). "On the subjective probability of compound events." *Organizational Behavior and Human Decision Processing 9*(3): 396–406.

Chapman, L. J., and J. P. Chapman. (1969). "Illusory correlation as an obstacle to the use of valid psychodiagnostic signs." *Journal of Abnormal Psychology 74*(3): 271–280.

Cohen, J., E. I. Chesnick, & D. Haran. (1972). "A confirmation of the intertial-PSI effect in sequential choice and decision." *British Journal of Psychology 63*(1): 41–46.

De Finetti, B. (1968). *International Encyclopedia of the Social Sciences*. D. E. Sills. New York, MacMillan. *12*: 496–504.

Edwards, W. (1968). In B. Kleinmuntz (Ed.), *Formal Representation of Human Judgment*. New York, Wiley.

Galbraith, R. C., and B. J. Underwood. (1973). "Perceived frequency of concrete and abstract words." *Memory Cognition 1*(1): 56–60.

Kahneman, D., and A. Tversky. (1972). "Subjective probability: a judgment of representativeness." *Cognitive Psychology 3*(3): 430–454.

Kahneman, D., and A. Tversky. (1973). "On the psychology of prediction." *Psychological Review 80*(4): 237–251.

Savage, L. J. (1954). *The Foudations of Statistics*. New York, Wiley.

Slovic, P., and S. Lichenstein. (1971). "Comparison of Bayesian and regression analysis approaches to the study of information processing in judgment." *Organizational Behavior and Human Decision Processing 6*(6): 649–744.

Staël von Holstein, C. A. (1971). "Two techniques for assessment of subjective probability judgments: An experimental study." *Acta Psychologica 35*(6): 478–494.

Tversky, A., and D. Kahneman. (1971). "Belief in law of small numbers." *Psychological Bulletin 76*(2): 105–110.

Tversky, A., and D. Kahneman. (1973). "Availability: A heuristic for judging frequency and probability." *Cognitive Psychology 5*(2): 207–232.

Winkler, R. L. (1967). "The quantification of judgment: Some methodological suggestions." *Journal of the American Statistical Association 62*(320): 1105–1120.

Chapter 26

Decision Making

Eldar Shafir and Amos Tversky

26.1 Introduction

Decisions about what to buy, whom to vote for, or where to live shape many aspects of our lives. The study of decision making is an interdisciplinary enterprise involving economics, political science, and psychology, as well as statistics and philosophy. One can distinguish two approaches to the analysis of decision making, the normative and the descriptive. The normative approach, which underlies much of economic analysis, assumes a rational decision maker, who has well-defined preferences that do not depend on the particular description of the options or on the specific methods for eliciting preference. This conception, which has come to be known as the rational theory of choice, is based primarily on a priori considerations rather than on experimental observation. As a consequence, it has a better claim as a normative account of how decisions ought to be made than as a descriptive theory of how decisions are actually made.

The descriptive approach to individual decision making is based on empirical observation and experimental studies of choice behavior. The experimental evidence indicates that people's choices are often at odds with the assumptions of the rational theory, and suggests some empirical generalizations that characterize people's choices. In this chapter we describe some selected findings and discuss several psychological principles that underlie the decision-making process. In the next section we address the psychological evaluation of gains and losses, and consider people's attitudes toward risk. Section 26.3 demonstrates that alternative descriptions of a decision problem can give rise to predictably different choices. Section 26.4 addresses the asymmetry between the evaluation of gains and losses, known as *loss aversion*. Section 26.5 demonstrates how alternative methods of eliciting people's preferences give rise to inconsistent decisions. In section 26.6 we address the role of conflict and show how preference among options is altered by the addition of new alternatives. The tension between descriptive and normative conceptions of decision making is addressed in the concluding section.

26.2 Risk and Value

Many decisions in the real world (such as investment, gambling, insurance) are risky in the sense that their outcomes are not known with certainty. To make

From chapter 3 in *An Invitation to Cognitive Science*, Vol. 3: *Thinking*, 2d ed., ed. E. E. Smith and D. N. Osherson (Cambridge, MA: MIT Press, 1995), 77–100. Reprinted with permission.

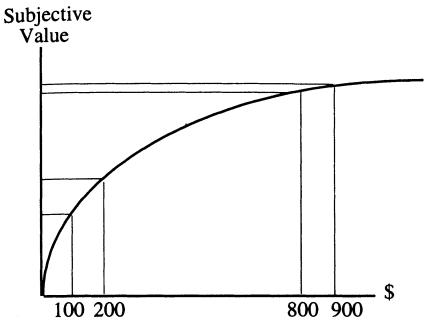


Figure 26.1
For gains, subjective value, or utility, is a concave function of money. A gain (or loss) of $100, for example, has a different subjective value depending on whether you have $100 or $800 to begin with.

such decisions, one has to consider two factors, the desirability of the potential outcomes and their probability of occurrence. Indeed, decision theory is concerned with the question of how these factors are, or should be, combined.

Consider a choice between a risky prospect that offers a 50 percent chance to win $200 (and a 50 percent chance to win nothing) and the alternative of receiving $100 for sure. Most people prefer the sure gain over the gamble, although the two prospects have the same expected value. The expected value of a gamble is a weighted average where each possible outcome is weighted by its probability of occurrence. The expected value of the gamble above is $.50 \times \$200 + .50 \times 0 = \$100$. A preference for a sure outcome over a risky prospect that has higher or equal expected value is called *risk averse*; a preference for a risky prospect over a sure outcome that has higher or equal expected value is called *risk seeking*.

As illustrated above, people tend to be risk averse when choosing between prospects with positive outcomes. This tendency toward risk aversion can be explained by appealing to the notion of diminishing sensitivity. Just as the impact of a candle is greater when it is brought into a dark room than into a room that is well lit, so the impact of an additional $100 is greater when it is added to a gain of $100 than when it is added to a gain of $800. This principle was first formalized by Daniel Bernoulli and Gabriel Cramer, who proposed early in the eighteenth century that subjective value, or utility, is a concave function of money, as illustrated in figure 26.1. (A function is concave if a line joining any two points on the curve lies entirely below the curve.) Notice that according to such a function the utility difference, $u(\$200) - u(\$100)$, is greater than the utility difference, $u(\$900) - u(\$800)$, though the dollar differences are the same.

Bernoulli and Cramer proposed that a person has a concave utility function that captures his or her subjective value for money, and that preferences should be described using expected utility instead of expected value. According to expected utility, the worth of a gamble offering a 50 percent chance to win $200 (and a 50 percent chance to win nothing) is $.50 \times u(\$200)$, where u is the person's utility function. (Assume that $u(0) = 0$.) As can be seen from figure 26.2, it follows from such a function that the subjective value attached to a gain of $100

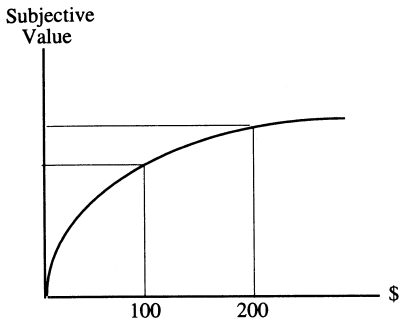


Figure 26.2
The subjective value curve can be used to illustrate risk behaviors. Here, the subjective value of a \$100 gain is seen to be more than $\frac{1}{2}$ the value of a \$200 gain, entailing preference for the "sure thing" \$100 gain described in the text.

is more than 50 percent of the value attached to a gain of \$200, which entails preference for the sure \$100 gain and, hence, risk aversion. Expected utility theory and the assumption of risk aversion play a central role in the standard economic analysis of choice between risky prospects.

Let us turn now to choice involving losses. Suppose you are forced to choose between a prospect that offers a 50 percent chance to lose \$200 (and a 50 percent chance to lose nothing) and the alternative of losing \$100 for sure. In this problem, most people reject the sure loss of \$100 and prefer to take an even chance at losing \$200 or nothing. Notice that, as in the choice above involving gains, the prospects have the same expected value. This preference for a risky prospect over a sure outcome that has the same expected value is an instance of risk seeking. Evidently, risk aversion does not always hold, in contrast to traditional economic analysis. In fact, except for prospects that involve very small probabilities, risk aversion is generally observed in choices involving gains, whereas risk seeking tends to hold in choices involving losses.

The combination of risk aversion for gains and risk seeking for losses can be explained by assuming that diminishing sensitivity applies to negative as well as to positive outcomes. Consequently, the subjective value function for losses is convex, as depicted in figure 26.3. (A function is convex if a line joining any two points on the curve lies entirely above the curve.) According to such a function, the worth of a gamble that offers a 50 percent chance to lose \$200 is greater (that is, less negative) than that of a sure loss of \$100. That is, $.50 \times u(-\$200) > u(-\$100)$. This result implies a risk-seeking preference for the gamble over the sure loss.

By conjoining figures 26.2 and 26.3, we obtain an S-shaped value function that is concave for gains and convex for losses, as illustrated in figure 26.4. This function forms part of a descriptive analysis of choice, known as *Prospect Theory*, which accounts for observed regularities in risky choice (Kahneman and Tversky 1979; Tversky and Kahneman 1992). The value function of Prospect Theory has three important properties: (1) it is defined on gains and losses rather than total wealth, (2) it is steeper for losses than for gains, and (3) it is concave for gains and convex for losses. The first property states that people

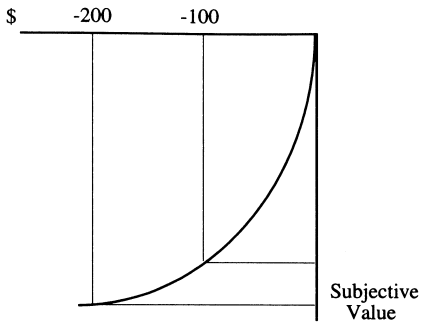


Figure 26.3
For losses, subjective value, or utility, is a convex function of money.

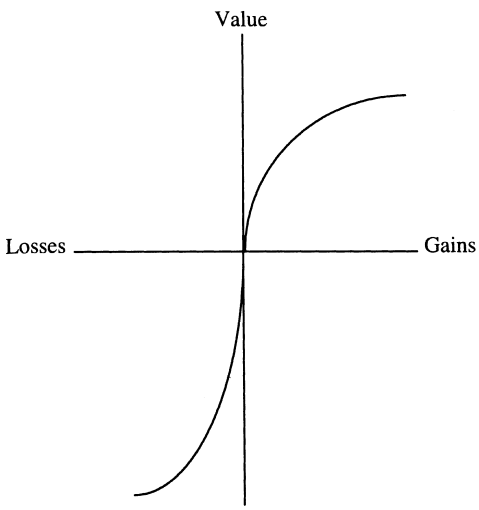


Figure 26.4
Under prospect theory, the concave gain function and convex loss function of figures 26.1 and 26.3 are combined.

normally treat outcomes as gains and losses defined relative to a neutral reference point, rather than in terms of total wealth, as we shall illustrate. The second property, called *loss aversion*, states that losses generally loom larger than corresponding gains. Thus, a loss of \$X is more aversive than a gain of \$X is attractive, which is implied by a function that is steeper for losses than for gains, that is, where $u(\$X) < -u(-\$X)$, as in figure 26.4.

The third property of the value function implies the risk attitudes described earlier: risk aversion in the domain of gains and risk seeking in the domain of losses. Although there is a presumption that people are entitled to their own values and each of the attitudes above seems unobjectionable on its own, the combination of the two leads to unacceptable consequences, as we shall show.

26.3 Framing Effects

Consider the following problems (Tversky and Kahneman 1986). The numbers in brackets indicate the percentage of respondents who chose each option. (The number of respondents in each problem is denoted N.)

Problem 1 ($N = 126$)
Assume yourself richer by \$300 than you are today.
You have to choose between
- a sure gain of \$100 [72%]
- a 50% chance to gain \$200 and a 50% chance to gain nothing [28%]

Problem 2 ($N = 128$)
Assume yourself richer by \$500 than you are today.
You have to choose between
- a sure loss of \$100 [36%]
- a 50% chance to lose nothing and a 50% chance to lose \$200 [64%]

In accord with the value function above, most subjects presented with problem 1, which is framed as a choice between gains, are risk averse, whereas most subjects presented with problem 2, which is framed as a choice between losses, are risk seeking. However, the two problems are essentially identical: When the initial payment of \$300 or \$500 is added to the respective outcomes, both problems amount to a choice between \$400 for sure and an even chance at \$300 or \$500. The different responses to problems 1 and 2 show that subjects did not combine the initial payment with the choice outcomes as required by normative analysis. As a consequence, the same choice problem framed in alternative ways led to systematically different choices. This result is called a *framing effect*.

The combination of risk aversion for gains and risk seeking for losses implied by the value function of figure 26.4 can also lead to violations of dominance, which is perhaps the simplest and most compelling principle of rational choice. The dominance principle states that if option B is better than option A on one attribute and at least as good as A on all the rest, then B should be chosen over A. For example, given a choice between

A: 25% chance to win \$240 and 75% chance to lose \$760
B: 25% chance to win \$250 and 75% chance to lose \$750

the dominance principle requires that the decision maker prefer option B to option A, because B offers the same chances of winning more than A and of losing less. Consider, in contrast, the following two choices, one involving gains and the other involving losses (Tversky and Kahneman 1981):

Problem 3 ($N = 150$)
Imagine that you face the following pair of concurrent decisions.
First examine both decisions, then indicate the options you prefer.

Decision (i). Choose between
C: a sure gain of \$240. [84%]
D: 25% chance to gain \$1,000 and 75% chance to gain nothing [16%]

Decision (ii). Choose between

E: a sure loss of \$750 [13%]

F: 75% chance to lose \$1,000 and 25% chance to lose nothing [87%]

Notice that the expected value of option D is $.25 \times \$1{,}000 = \$250$, whereas the expected value of option F is $.75 \times -\$1{,}000 = -\$750$. Hence, as the data show, the majority choice in decision (i) is risk averse, and the majority choice in decision (ii) is risk seeking, as predicted by the value function. As it turns out, 73 percent of the subjects chose a combination of the two most popular options, C and F, and only 3 percent of the subjects chose a combination of the two least popular prospects, D and E. Simple calculation, however, shows that the combination of C and F yields prospect A above, whereas the combination of D and E yields prospect B. Thus, a great majority of subjects violated dominance and selected an inferior combination of prospects. In contrast, when subjects were presented with a direct choice between A and B, everybody naturally chose the dominant option B. Thus, the principle of dominance is obeyed when its application is transparent, but is often violated when it is not. In particular, the demonstration above shows that the tendency to evaluate prospects in isolation, combined with the common risk attitudes captured by figure 26.4, can lead to the selection of a dominated option.

The effects of framing and the characteristics of the value function are not limited to monetary outcomes, as demonstrated by the following choices between health policies involving human life (Tversky and Kahneman 1981):

Problem 4 ($N = 152$)

Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

If Program A is adopted, 200 people will be saved. [72%]

If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. [28%]

Notice that both programs have the same expected value in terms of human lives. Because saving people is perceived as a "gain," the majority of subjects made the risk-averse choice of saving 200 people for sure over the chance of saving either 600 people or no one. A second group of subjects was given the same cover story with these descriptions of the alternative programs:

Problem 5 ($N = 155$)

If Program C is adopted, 400 people will die. [22%]

If Program D is adopted, there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die. [78%]

Here the outcomes of the two programs are described in terms of lives lost. Accordingly, the majority of subjects made the risk-seeking choice, avoiding the sure loss of 400 lives in favor of the chance to save either all 600 or no one. Subjects again exhibited the familiar pattern of risk aversion in the domain of gains and risk seeking in losses. However, problems 4 and 5 present the same

options. In particular, programs A and B are identical, respectively, to programs C and D. They differ only in that the former are framed in terms of number of lives saved, whereas the latter are framed in terms of lives lost.

An essential element in the rational theory of choice is the requirement, known as *description invariance*, that equivalent representations of a choice problem should yield the same preferences. That is, an individual's preference between options should not depend on the manner in which they are described, provided the descriptions convey the same information. The majority preferences expressed in problems 4 and 5, however, violate the principle of description invariance and show that framing the same problem in terms of gains or in terms of losses gives rise to predictably different choices.

Framing effects are pervasive and are often observed even when the same respondents answer both versions of a problem. Furthermore, they are found in the choices of both naive and sophisticated respondents. For example, experienced physicians made markedly different choices between two alternative treatments for lung cancer—surgery and radiation therapy—depending on whether the outcomes of these treatments were described in terms of mortality rates or in terms of survival rates. Surprisingly, the physicians were just as susceptible to the effect of framing as were graduate students or clinic patients (McNeil, Pauker, Sox, and Tversky 1982).

The effectiveness of framing manipulations suggests that people tend to adopt the frame presented in a problem and evaluate the outcomes in terms of that frame. Thus, depending on whether a problem is described in terms of gains or losses, people are likely to exhibit risk-averse or risk-seeking behaviors. An interesting class of framing effects arises in the evaluations of economic transactions that occur in times of inflation.

In one study (Shafir, Diamond, and Tversky 1994), subjects were asked to imagine that they worked for a company that produced computers in Singapore, and had to sign a contract for the local sale of new computers in that country. The computers, currently selling for $1,000 apiece, were to be delivered and paid for a year later. By that time, due to inflation, all prices, including production costs and computer prices, were expected to increase by about 20 percent. Subjects had to choose between contract A: selling the computers a year later for the predetermined price of $1,200 (that is, 20 percent higher than the current price), and contract B: selling the computers a year later for the going price at that time. For one group of subjects the options were described relative to the predetermined price of $1,200. In this frame, contract A appears riskless because the computers are guaranteed to sell for $1,200 no matter what, whereas contract B appears risky because the computers' future price will be less than $1,200 if inflation is low, and more than $1,200 if inflation is high. A second group of subjects were presented with the same alternatives described relative to the computers' expected future price. Here, contract B appears riskless because the computers will be sold next year for their actual price then, regardless of the rate of inflation. Contract A, on the other hand, appears risky: the computers are to be sold for $1,200, which may be more than they are worth if inflation is lower than the anticipated 20 percent, and less than they are worth if inflation exceeds 20 percent. Because of loss aversion, the contract that appeared riskless in each frame was relatively more attractive than

the one that appeared risky. Thus, contract A was chosen more often in the former case, when it was framed as riskless, than in the latter, when it was framed as risky.

26.4 Loss Aversion

One of the basic observations regarding people's reaction to outcomes is that losses appear larger than corresponding gains. This asymmetry in the evaluation of positive and negative outcomes is called *loss aversion*. Loss aversion gives rise to a value function that is steeper in the negative than in the positive domain, as in figure 26.4. An immediate implication of loss aversion is that people will not accept an even chance to win or lose $X, because the loss of $X is more aversive than the gain of $X is attractive. Indeed, people are generally willing to accept an even-chance prospect only when the gain is substantially greater than the loss. Many people, for example, reject a 50–50 chance to win $200 or lose $100, even though the gain is twice as large as the loss (Tversky and Shafir 1992a).

The example above illustrates loss aversion in decisions involving risky prospects. The principle of loss aversion applies with equal force to riskless choice, between options that can be obtained for certain (Tversky and Kahneman 1991). It entails that the loss of utility associated with giving up a good that is in our possession is generally greater than the utility gain associated with obtaining that good. An instructive demonstration of this effect is provided in an experiment involving the selling of mugs (Kahneman, Knetsch, and Thaler 1990). A class is divided into two groups. Some participants, called *sellers*, are given a decorated mug that they can keep, and are asked to indicate the lowest price for which they would be willing to sell the mug. A second group, called *choosers*, are asked to indicate the amount of money that they would find as attractive as the mug. Subjects in both groups are told that, after they state their price, an official market price $X will be revealed and that each subject will end up with a mug if his or her asking price exceeds $X, or with $X if it is more than the subject's asking price.

Notice that the choosers and the sellers are facing precisely the same decision problem: they will all end up with either some money or a mug, and in effect need to decide how much money they will be willing to take in place of the mug. Hence, standard economic analysis predicts identical asking prices for the two groups. The two groups, however, evaluate the mug from different perspectives: the choosers compare receiving a mug to receiving a sum of money, whereas the sellers compare retaining the mug to giving up the mug in exchange for money. Thus, the mug is evaluated as a potential gain by the choosers and as a loss by the sellers. Consequently, loss aversion, the notion that losses loom larger than corresponding gains, predicts that the sellers will price the mug higher than the choosers. This prediction was confirmed by the data: the median price of the sellers ($7.12) was more than twice as large as the median price for the choosers ($3.12). The difference between these prices reflects an endowment effect, which was produced, instantaneously it seems, by endowing individuals with a mug.

A closely related manifestation of loss aversion is a general reluctance to trade, which is illustrated in the following study (Knetsch 1989). Subjects were

divided into two groups: half the subjects were given a decorated mug, and the others were given a large bar of Swiss chocolate. Later, each subject was shown the alternative gift, and offered the opportunity to trade the gift they had received for the other. Because the initial allocation of gifts was arbitrary and the transaction was costless, economic theory predicts that about half the subjects should exchange their gifts. On the other hand, if losses loom larger than gains, then most participants will be reluctant to give up the gift in their possession (a loss) in order to obtain the other (a gain). Indeed, only 10 percent of the participants chose to trade their gifts. This result contrasts sharply with the 50 percent predicted by the standard economic analysis, in which the value of a good does not change when it becomes part of one's endowment.

More generally, loss aversion entails a strong tendency to maintain the status quo, because the disadvantages of departing from it loom larger than the advantages of its alternative. This phenomenon has been demonstrated in several experiments (Samuelson and Zeckhauser 1988). For example, subjects were given this problem: "You have inherited a large sum of money from your great uncle. You are considering different portfolios. Your choices are to invest in (1) a moderate-risk company, (2) a high-risk company, (3) treasury bills, (4) municipal bonds." Four groups of subjects were presented with the same problem, but with one of the four options designated as the status quo. One version, for example, included the statement: "A significant portion of the portfolio you inherited is invested in a moderate-risk company." The data show that designating a particular option as the status quo greatly increased the tendency to choose it (even though transaction costs were said to be insignificant). Although all four groups chose among the same options, subjects tended to stick with the option in which they were already invested.

A striking framing effect that relies on people's tendency to maintain the status quo has been observed in the context of real-world insurance decisions. New Jersey and Pennsylvania have recently introduced the option of a limited right to sue, which entitles automobile drivers to lower insurance rates. The two states differ, however, in what they offer consumers as the default option. New Jersey motorists have to acquire the full right to sue (transaction costs are minimal: one need only sign), whereas in Pennsylvania the full right to sue is the default. When offered the choice, only about 20 percent of New Jersey drivers chose to acquire the full right to sue, but approximately 75 percent of Pennsylvania drivers chose to retain it. The difference in adoption rates due to the different frames had financial repercussions that are estimated at around $200 million (Johnson, Hershey, Meszaros, and Kunreuther 1993).

Recall that loss aversion gives rise to a value function with a steeper slope in the negative than in the positive domain. Beyond the reluctance to depart from the status quo, this result implies that the same difference between two options will be given greater weight when it is viewed as a difference between two disadvantages, or losses (relative to a reference point) than when it is viewed as a difference between two advantages, or gains. This prediction is demonstrated in a study in which subjects compare a combination of a small gain and a small loss with a combination of a larger gain and a larger loss (Tversky and Kahneman 1991). Subjects are asked to suppose that they are looking for employment while their present training job is ending. They are asked to consider two

alternative jobs that are like their present job in all respects except for the amount of social contact and the daily commuting time. The relevant information is summarized in the accompanying table. Subjects are divided into two groups: one group is told that they presently hold job A, the second group is told they presently hold job B. Both groups are then asked to choose between job X and job Y. Because their current jobs are said to be ending, maintaining the status quo is not an option.

	Social Contact	*Daily Travel Time*
Present Job A	isolated for long stretches	10 min.
Job X	limited contact with others	20 min.
Job Y	moderately sociable	60 min.
Present Job B	much pleasant social interaction	80 min.

Notice that both X and Y are better than A and worse than B with respect to social contact, and both are worse than A and better than B in terms of commuting time. According to standard economic analysis, the choice between X and Y should not depend on the decision maker's current reference point. On the other hand, if subjects treat their present job as a reference point and if disadvantages relative to this reference point loom larger than corresponding advantages, then subjects are more likely to choose the job with the smaller disadvantage relative to their current job. Thus, subjects who currently hold job A are expected to favor job X, whereas subjects who currently hold job B are expected to favor job Y. The data confirm this expectation: more than two-thirds of subjects in each group chose the predicted option.

Loss aversion, or the asymmetry between the evaluation of gains and losses, emerges as an important empirical generalization that has implications for a wide range of decisions. It promotes stability rather than change by inducing people to maintain their current position. A loss-averse individual at position X would be reluctant to switch to position Y, even though, were she at position Y, she would be reluctant to switch to X. Along these lines, the reluctance to change induced by loss aversion can hinder the negotiated resolution of disputes. If each side to a dispute evaluates the opponent's concessions as gains and its own concessions as losses, then agreement will be hard to reach because each side will perceive itself as relinquishing more than it stands to gain. A skillful mediator may facilitate agreement by framing concessions as bargaining chips rather than as losses.

26.5 Eliciting Preference

Preferences can be elicited by different methods. People can be asked to indicate which option they prefer; alternatively, they can be asked to price each option by stating the amount of money that is as valuable to them as that option. A standard assumption, known as *procedure invariance*, demands that logically equivalent elicitation procedures should give rise to the same preference order. Thus, if one option is chosen over another, it is also expected to be priced higher. Procedure invariance is essential for interpreting both psychological and physical measurement. For example, the ordering of physical objects with

respect to mass can be established either by placing each object separately on a scale, or by placing both objects on two sides of a pan balance. Procedure invariance requires that the two methods yield the same ordering, within the limit of measurement error. Analogously, the rational theory of choice assumes that an individual has a well-defined preference order that can be elicited either by choice or by pricing. These alternative methods of elicitation, in turn, should give rise to the same ordering of options.

26.5.1 Compatibility Effects

Despite its appeal as an abstract principle, people sometimes violate procedure invariance. For example, people often choose one bet over another, but price the second bet above the first. In one study, subjects were presented with two prospects of similar expected value. One prospect, the H bet, offered a high probability to win a relatively small payoff (for example, 8 chances in 9 to win $4) whereas the other prospect, the L bet, offered a low probability to win a larger payoff (for example, a 1 in 9 chance to win $40). When asked to choose between these prospects, most subjects chose the H bet over the L bet. Subjects were also asked, on another occasion, to price each prospect by indicating the smallest amount of money for which they would be willing to sell this prospect. Here, most subjects assigned a higher price to the L bet than to the H bet. One recent study that used this pair of bets observed that 71 percent of the subjects chose the H bet, and 67 percent priced L above H (Tversky, Slovic, and Kahneman 1990). This phenomenon, called *preference reversal*, has been observed in numerous experiments using a variety of prospects and incentive schemes. It has also been observed among professional gamblers in a Las Vegas casino (Slovic and Lichtenstein 1983).

What is the cause of preference reversal? Why do people assign a higher monetary value to the low-probability bet, but choose the high-probability bet more often? It appears that the major cause of preference reversal is a differential weighting of probability and payoff in choice and pricing, induced by the required response. In particular, experimental evidence indicates that an attribute of an option is given more weight when it is compatible with the response format than when it is not (Tversky, Sattath, and Slovic 1988). This account suggests that because the price that the subject assigns to a bet is expressed in dollars, the payoffs of the bet, which are also expressed in dollars, will be weighted more heavily in pricing than in choice. As a consequence, the L bet (which has the higher payoff) is evaluated more favorably in pricing than in choice, which can give rise to preference reversals. This account has been supported by the observation that the incidence of preference reversals was greatly reduced for bets involving nonmonetary outcomes, such as a free dinner at a local restaurant, where the outcomes and the subjects' prices are no longer expressed in the same units and are therefore less compatible (Slovic, Griffin, and Tversky 1990).

The compatibility hypothesis does not depend on the presence of risk. Indeed, it predicts a similar discrepancy between choice and pricing in the context of riskless options that have a monetary component. Consider a long-term prospect L, which pays $2,500 five years from now, and a short-term prospect S, which pays $1,600 in one and a half years. Subjects were invited to choose

between L and S and to price both prospects by stating the smallest immediate cash payment for which they would be willing to exchange each prospect (Tversky, Slovic, and Kahneman 1990). Because the payoffs and the prices again are expressed in the same units, compatibility suggests that the long-term prospect (offering the higher payoff) will be overvalued in pricing relative to choice. In accord with this hypothesis, subjects chose the short-term prospect 74 percent of the time but priced the long-term prospect above the short-term prospect 75 percent of the time. These observations indicate that different methods of elicitation (for example, choice and pricing) can induce different weightings of attributes that in turn give rise to preference reversals.

26.5.2 Relative Prominence

Another psychological mechanism that leads to violations of procedure invariance involves the notion of relative prominence. In many cases, people agree that one attribute (for instance, safety) is more important than another (such as cost). Although the interpretation of such claims is not entirely clear, there is evidence that the attribute that is judged more important looms larger in choice than in pricing (Tversky, Sattath, and Slovic 1988). This is the *prominence hypothesis*. To illustrate this notion, consider two programs designed to reduce the number of fatalities due to traffic accidents, characterized by the expected reduction in the number of casualties and an estimated cost. Because human lives are regarded as more important than money, the prominence hypothesis predicts that this dimension will be given more weight in choice than in pricing. When given a choice between programs X and Y (see accompanying table), the great majority of respondents favored X, the more expensive program that saves more lives.

	Expected Number of Casualties	*Cost*
Program X	500	$55 million
Program Y	570	$12 million

However, when the cost of one of the programs is removed and subjects are asked to determine the missing cost so as to make the two programs equally attractive, nearly all subjects assign values that imply a preference for Y, the less expensive program that saves fewer lives. For example, when the cost of program X is removed, the median estimate of the missing cost that renders the two programs equally attractive is $40 million. This choice implies that at $55 million, program X should not be chosen over program Y, contrary to the aforementioned choice. Thus, the prominent attribute (saving lives) dominates the choice but not the pricing. This discrepancy suggests that different public policies will be supported depending on whether people are asked which policy they prefer or how much, in their opinion, each policy ought to cost.

Further applications of the prominence hypothesis were reported in a study of people's response to environmental problems (Kahneman and Ritov 1993). Several pairs of issues were selected, where one issue involves human health or safety and the other protection of the environment. Each issue includes a brief statement of a problem, along with a suggested form of intervention, as illustrated.

Problem: Skin cancer from sun exposure is common among farm workers.
Intervention: Support free medical checkups for threatened groups.

Problem: Several Australian mammal species are nearly wiped out by hunters.
Intervention: Contribute to a fund to provide safe breeding areas for these species.

One group of subjects was asked to choose which of the two interventions they would rather support; a second group of subjects was presented with one issue at a time and asked to determine the largest amount they would be willing to pay for the respective intervention. Because the treatment of cancer in human beings is generally viewed as more important than the protection of Australian mammals, the prominence hypothesis predicts that the former will receive greater support in direct choice than in independent evaluation. This prediction was confirmed. When asked to evaluate each intervention separately, subjects, who might have been moved by these animals' plight, were willing to pay more, on average, for safe breeding of Australian mammals than for free checkups for skin cancer. When faced with a direct choice between these options, however, most subjects favored free checkups for humans over safe breeding for mammals. Thus, people may evaluate one alternative more positively than another when each is evaluated independently, but then reverse their evaluation when the alternatives are directly compared, which accentuates the prominent attribute.

26.5.3 Weighing Pros and Cons

Consider having to choose one of two options or, alternatively, having to reject one of two options. Under the assumption of procedure invariance, the two tasks are interchangeable. In binary choice it should not matter whether people are asked which option they prefer, or which they would reject: if people prefer the first they should reject the second, and vice versa. In line with the notion of compatibility, however, we may expect that the positive features of options (their pros) will loom larger when choosing, whereas the negative features of options (their cons) will be weighted more heavily when rejecting. It is natural to select an option because of its positive features, and to reject an option because of its negative features.

This account leads to the following prediction: Imagine two options, an "enriched" option, with many positive and many negative features, and an "impoverished" option, with few positive and few negative features. If positive features are weighed more heavily when choosing than when rejecting and negative features are weighed more heavily when rejecting than when choosing, then an enriched option could be both chosen and rejected more frequently than an impoverished option. Consider, for example, the following problem, which was presented to subjects in two versions that differed only in the bracketed questions (Shafir 1993). Half the subjects received one version, the other half received the other.

Problem 6 ($N = 170$)
Imagine that you serve on the jury of an only-child sole-custody case following a relatively messy divorce. The facts of the case are

complicated by ambiguous economic, social, and emotional considerations, and you decide to base your decision entirely on the following few observations. [To which parent would you award sole custody of the child?/To which parent would you deny sole custody of the child?]

		Award	*Deny*
Parent A	average income		
	average health		
	average working hours		
	reasonable rapport with the child	[36%]	[45%]
	relatively stable social life		
Parent B	above-average income		
	very close relationship with the child	[64%]	[55%]
	extremely active social life		
	lots of work-related travel		
	minor health problems		

Parent A, the impoverished option, is quite plain—with no striking positive or negative features. There are no particularly compelling reasons to award or deny this parent custody of the child. Parent B, the enriched option, on the other hand, has good reasons to be awarded custody (a very close relationship with the child and a good income), but also good reasons to be denied sole custody (health problems and extensive absences due to travel). To the right of the options are the percentages of subjects who chose to award and to deny custody to each of the parents. Parent B is the majority choice both for being awarded custody of the child and for being denied it, presumably because this parent provides more compelling reasons both to be awarded as well as denied child custody. As a result, the child is significantly more likely to end up with parent B when we ask whom to award custody to than when we contemplate whom to deny. This discrepancy represents another violation of procedure invariance, in which two logically equivalent tasks give rise to predictably different choices.

26.6 Choice under Conflict

The rational theory of choice assumes that each alternative has a utility or subjective value for the decision maker. Given a set of options, the decision maker selects the alternative with the highest value. This principle of value maximization is routinely assumed in analyzing consumer choice. It implies that the preference between options cannot be reversed by the addition of new alternatives. If you prefer salmon to steak, for example, you should not select steak from a larger menu that includes salmon, unless, of course, other entrées provide some information about the quality of the steak or the salmon. Thus, a nonpreferred option cannot be made preferred by introducing new alternatives. Consequently, the "market share" of an option (that is, the proportion of people

who select it) cannot be increased when new options are added. In particular, the proportion of people who choose the option to defer decision should not increase when additional alternatives become available.

Despite the simplicity and intuitive appeal of the principle above, there is evidence that people's preference between two options can depend on the presence or absence of a third alternative. The introduction of a third option can make the decision easier or harder to resolve and thus can affect preference and increase the tendency to defer choice. The making of decisions often creates conflict: we are not sure how to trade off one attribute relative to another or which option would benefit us most. When people are offered a single attractive option, there is little conflict and choice is easy; however, when two or more attractive options are available, each with its advantages and disadvantages, people often experience conflict, which may compel them to delay decision, maintain the status quo, or seek additional information.

The economist Thomas Schelling tells of an occasion on which he had decided to buy an encyclopedia for his children, and was presented at a bookstore with two attractive options. Finding it difficult to choose between them, he ended up buying neither, although had only one encyclopedia been available, he would have happily bought it. More generally, there are situations in which people prefer each of the available alternatives over the status quo but do not have a compelling reason for choosing among the alternatives and, as a result, defer the decision, perhaps indefinitely.

This phenomenon is demonstrated by this pair of problems, which were presented to two groups of students (Tversky and Shafir 1992b).

> *Problem 7* ($N = 121$), *Low Conflict*
> Suppose you are considering buying a compact disc (CD) player, and have not yet decided what model to buy. You pass by a store that is having a one-day clearance sale. They offer a popular SONY player for just $99, well below the list price. Do you?
>
> y. buy the SONY player [66%]
> z. wait until you learn more about the various models [34%]

> *Problem 8* ($N = 124$), *High Conflict*
> Suppose you are considering buying a compact disc (CD) player, and have not yet decided what model to buy. You pass by a store that is having a one-day clearance sale. They offer a popular SONY player for just $99, and a top-of-the-line AIWA player for just $169, both well below the list price. Do you?
>
> x. buy the AIWA player [27%]
> y. buy the SONY player [27%]
> z. wait until you learn more about the various models [46%]

The results indicate that people are more likely to buy a CD player in the former, *low conflict*, condition than in the latter, *high conflict*, situation. Both products—the AIWA and the SONY—seem attractive, both are well priced, and both are on a one-day sale. The decision maker needs to determine whether she is better off with a cheaper, popular product, or with a more ex-

pensive and sophisticated one. This conflict is not easy to resolve, and compels many subjects to put off the purchase until they learn more about the various products. On the other hand, when the SONY alone is available, there are compelling arguments for its purchase: it is a popular player, it is very well priced, and it is on sale for one day only. In this situation, a greater majority of subjects decide to opt for the CD player rather than delay the purchase.

Adding a competing alternative in the preceding example increased the tendency to delay decision. Adding an option can also have the opposite effect, as illustrated in this problem, in which the original AIWA player was replaced by an inferior model.

> *Problem 9* ($N = 62$), *Dominance*
> Suppose you are considering buying a compact disc (CD) player, and have not yet decided what model to buy. You pass by a store that is having a one-day clearance sale. They offer a popular SONY player for just $99, well below the list price, and an inferior AIWA player for the regular list price of $105. Do you?
>
> x′. buy the AIWA player [3%]
> y. buy the SONY player [73%]
> z. wait until you learn more about the various models [24%]

In this version, the AIWA player is dominated by the SONY: it is inferior in quality and costs more. Thus, the presence of the AIWA does not detract from the reasons for buying the SONY, it actually supplements them: the SONY is well priced, it is on sale for one day only, *and* it is clearly better than its competitor. As a result, in the presence of the inferior AIWA, the SONY is chosen more often. More generally, adding a dominated alternative tends to increase the market share of the dominating option (Huber, Payne, and Puto 1982), contrary to the prediction of value maximization.

In the scenario above, the added options (the superior CD player in one case and the inferior player in the other) may have conveyed some information about the consumer's chances of finding a better deal. This interpretation does not apply to the following demonstrations, in which there is no opportunity to learn about the options, and the decision cannot be delayed. One group of subjects ($N = 106$) was offered a choice between $6 and an elegant Cross pen. The pen was selected by 36 percent of the subjects, and the remaining 64 percent chose the cash. A second group ($N = 115$) was given a choice among three options: $6 in cash, the same Cross pen, and a second pen that was distinctly less attractive. Only 2 percent of the subjects chose the less attractive pen, but its presence increased the percentage of subjects who chose the Cross pen from 36 percent to 46 percent (Simonson and Tversky 1992). Students of marketing recount many instances of the phenomenon above in the marketplace. A common tactic used to induce consumers to purchase a given product is to introduce an inferior option that renders the product in question more attractive. For example, Williams-Sonoma, a mail-order and retail business located in San Francisco, used to offer a bread-baking appliance priced at $275. They then added a second bread-baking appliance, very similar to the first except that it was larger but could not bake whole-wheat bread. The new item was priced at $429, more than 50 percent higher than the original appliance. Not surprisingly,

Williams-Sonoma did not sell many units of the new item, but the sales of the less-expensive appliance almost doubled.

The effect of added alternatives is not limited to decisions made by consumers. In one study (Redelmeier and Shafir 1995), 287 experienced physicians were presented with a description of a hypothetical patient suffering from chronic hip pain and about to be referred to orthopedics. Half the physicians were presented with a choice of whether or not to assign this patient a particular medication (Motrin); the other half were presented with two alternative medications (Motrin and Feldene). The proportion of physicians who refrained from assigning any new medication was 53 percent in the former group and 72 percent in the latter. Thus, the availability of a second medication reduced the tendency to assign either. Evidently, the difficulty of deciding which of the two medications was preferable led many physicians to avoid medication altogether.

The experimental evidence shows that, contrary to the principle of value maximization, the availability of additional alternatives can increase conflict and lead the decision maker to maintain the status quo, avoid the decision, or postpone it indefinitely. It is difficult to overestimate the significance of the tendency to delay decision. Many things never get done, not because one has chosen not to do them, but because the person has chosen not to do them *now*. The following demonstration illustrates this point. Students were offered $5 for answering and returning an assigned questionnaire by a given date. One group was given 5 days to complete the questionnaire, a second group was given 3 weeks, and a third group was given no definite deadline. The corresponding rates of return were 60 percent, 42 percent, and 25 percent. Thus, the more time students had to complete the task, the less likely they were to do it. Just as adding a second drug reduces the tendency to administer medication, so too can extending time reduce the likelihood of completing an assignment.

26.7 Discussion

In this chapter we have applied a number of psychological principles to the analysis of individual decision making. We have invoked the notion of diminishing sensitivity to derive the shape of the value function, which reflects people's evaluation of gains and losses. This function accounts for common observations of risk aversion in the domain of gains and risk seeking in the domain of losses. Because the same outcomes can sometimes be described either as gains or as losses, alternative framings of a decision problem can give rise to predictably different choices. We have also considered the principle of loss aversion, according to which losses have a greater impact than the corresponding gains. Loss aversion accounts for a wide range of findings, notably the reluctance to depart from the status quo.

Additional psychological principles were introduced to account for elicitation effects. We suggested that different attributes of options are weighted differently in choice and in pricing, and we invoked the notions of compatibility and prominence to explain the discrepancy between these procedures. Finally, we have appealed to considerations of conflict, or choice difficulty, to explain some effects of the addition of options and the tendency to defer decision.

The psychological principles discussed in this chapter do not form a unified theory, comparable to the rational theory of choice. However, they help explain a wide range of empirical findings that are incompatible with the rational theory. Recall that this theory assumes consistent preferences that satisfy description and procedure invariance. In contrast, the experimental evidence suggests that preferences are actually constructed, not merely revealed, in the elicitation process, and that these constructions depend on the framing of the problem, the method of elicitation, and the available set of options.

We have suggested that the rational theory of choice provides a better account of people's normative intuitions than of their actual behavior. When confronted with the fact that their choices violate dominance or description invariance, people typically wish to modify their behavior to conform with these principles of rationality. Evidently, people's choices are often at variance with their own normative intuitions. The tension between normative and descriptive theories of choice is analogous to the tension between normative and descriptive theories of ethics. A normative ethical account is concerned with the principles that underlie moral judgments. A descriptive ethical account, on the other hand, is concerned with actual human conduct. Both enterprises are essentially empirical; the first addresses people's intuitions, whereas the second focuses on their actual behavior. The two analyses, of course, are interrelated but they do not coincide. For example, people generally agree that one should abstain from lying and contribute to worthy causes, despite the fact they do not always do so. Similarly, people tend to accept the normative force of dominance and description invariance, even though these are often violated in their actual choices. Although the distinction between the normative and descriptive accounts is obvious in the study of ethics, it is somewhat controversial in the study of decision making. This difference may be due to the fact that it is easier to understand violations of ethical norms that stem from self-interest or lack of self-control, than violations of rational norms that stem from the nature of cognitive operations.

Suggestions for Further Reading

Elementary introductions to the field of behavioral decision theory include Bazerman (1992), Dawes (1988), Hogarth (1987), and Yates (1990). von Winterfeldt and Edwards (1986) is an introduction with more of an applied perspective, covering an area known as decision analysis. Thaler (1992) focuses on the role of behavioral theory in interpreting numerous economic anomalies. Shafir, Simonson, and Tversky (1993) consider the role of reasons in the making of decisions. For collections of primary articles relating behavioral decision theory to various domains of inquiry, ranging from economics and the law to engineering and philosophy, see Arkes and Hammond (1986), and Bell, Raiffa, and Tversky (1988). Recent reviews of the field are provided by Camerer (1995), Payne, Bettman, and Johnson (1992), and Slovic, Lichtenstein, and Fischhoff (1988).

Acknowledgments

Preparation of this chapter was supported by US Public Health Service Grant No. 1-R29-MH46885 from the National Institute of Mental Health, by Grant No. SBR-9408684 from the National Science Foundation, and by a grant from the Russell Sage Foundation. It was completed while the authors were Fellows at the Institute for Advanced Studies and the Center for Rationality and Interactive Decision Theory of The Hebrew University.

References

Arkes, H. R., and K. R. Hammond, eds. (1986). *Judgment and decision making*. New York: Cambridge University Press.

Bazerman, M. H. (1992). *Judgment in managerial decision making*. 2nd ed. New York: Wiley.

Bell, D. E., H. Raiffa, and A. Tversky, eds. (1988). *Decision making: Descriptive, normative, and prescriptive interactions*. New York: Cambridge University Press.

Camerer, C. F. (1995). Individual decision making. In J. H. Kagel and A. E. Roth, eds., *Handbook of experimental economics*. Princeton, NJ: Princeton University Press.

Dawes, R. M. (1988). *Rational choice in an uncertain world*. New York: Harcourt Brace Jovanovich.

Hogarth, R. M. (1987). *Judgment and choice*. 2nd ed. New York: Wiley.

Huber, J., J. W. Payne, and C. Puto (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research* 9, 90–98.

Johnson, E. J., J. Hershey, J. Meszaros, and H. Kunreuther (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty* 7, 35–51.

Kahneman, D., J. L. Knetsch, and R. Thaler (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy* 98, 6, 1325–1348.

Kahneman, D., and I. Ritov (1993). Determinants of stated willingness to pay for public goods: A study in the headline method. Working paper, University of California, Berkeley.

Kahneman, D., and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47, 263–291.

Knetsch, J. I. (1989). The endowment effect and evidence of nonreversible indifference curves. *American Economic Review* 79, 1277–1284.

McNeil, B. J., S. G. Pauker, H. C. Sox, and A. Tversky (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine* 306, 1259–1262.

Payne, J. W., J. R. Bettman, and E. J. Johnson (1992). Behavioral decision research: A constructive process perspective. *Annual Review of Psychology* 43, 87–131.

Redelmeier, D., and E. Shafir (1995). Medical decision making in situations that offer multiple alternatives. *Journal of the American Medical Association* 273, 4, 302–305.

Samuelson, W., and R. Zeckhauser (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty* 1, 7–59.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory and Cognition* 21, 4, 546–556.

Shafir, E., P. Diamond, and A. Tversky (1994). On money illusion. Manuscript, Princeton University.

Shafir, E., I. Simonson, and A. Tversky (1993). Reason-based choice. *Cognition* 49, 2, 11–36.

Simonson, I., and A. Tversky (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research* 29, 281–295.

Slovic, P., D. Griffin, and A. Tversky (1990). Compatibility effects in judgment and choice. In R. Hogarth, ed., *Insights in decision making: Theory and applications*, pp. 5–27. Chicago: University of Chicago Press.

Slovic, P., and S. Lichtenstein (1983). Preference reversals: A broader perspective. *American Economic Review* 73, 596–605.

Slovic, P., S. Lichtenstein, and B. Fischhoff (1988). Decision making. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, and R. D. Luce, eds., *Stevens' handbook of experimental psychology*. 2nd ed. New York: Wiley.

Thaler, R. (1992). *The winner's curse: Paradoxes and anomalies of economic life*. New York: The Free Press.

Tversky, A., and D. Kahneman (1981). The framing of decisions and the psychology of choice. *Science* 211, 453–458.

Tversky, A., and D. Kahneman (1986). Rational choice and the framing of decisions. *Journal of Business* 59, 4, pt. 2, 251–278.

Tversky, A., and D. Kahneman (1991). Loss aversion in riskless choice: A reference dependent model. *Quarterly Journal of Economics* (November), 1039–1061.

Tversky, A., and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5, 297–323.

Tversky, A., S. Sattath, and P. Slovic (1988). Contingent weighting in judgment and choice. *Psychological Review 95*, 3, 371–384.

Tversky, A., and E. Shafir (1992a). The disjunction effect in choice under uncertainty. *Psychological Science* 3, 5, 305–309.

Tversky, A., and E. Shafir (1992b). Choice under conflict: The dynamics of deferred decision. *Psychological Science* 3, 6, 358–361.

Tversky, A., P. Slovic, and D. Kahneman (1990). The causes of preference reversal. *American Economic Review* 80, 204–217.

von Winterfeldt, D., and W. Edwards (1986). *Decision analysis and behavioral research*. Cambridge: Cambridge University Press.

Yates, J. F. (1990). *Judgment and decision making*. Englewood Cliffs, NJ: Prentice-Hall.

Chapter 27

For Those Condemned to Study the Past: Heuristics and Biases in Hindsight

Baruch Fischhoff

Benson (1972) has identified four reasons for studying the past: to entertain, to create a group (or national) identity, to reveal the extent of human possibility, and to develop systematic knowledge about our world, knowledge that may eventually improve our ability to predict and control. On a conscious level, at least, we behavioral scientists restrict ourselves to the last motive. In its pursuit, we do case studies, program evaluations, and literature reviews. We even conduct experiments, creating artificial histories upon which we can perform our postmortems.

Three basic questions seem to arise in our retrospections: (a) Are there patterns upon which we can capitalize so as to make ourselves wiser in the future? (b) Are there instances of folly in which we can identify mistakes to avoid? (c) Are we really condemned to repeat the past if we do not study it? That is, do we really learn anything by looking backward?

Whatever the question we are asking, it is generally assumed that the past will readily reveal the answers it holds. Of hindsight and foresight, the latter appears as the troublesome perspective. One can explain and understand any old event if an appropriate effort is applied. Prediction, however, is acknowledged to be rather more tricky. The present essay investigates this presumption by taking a closer look at some archetypal attempts to tap the past. Perhaps its most general conclusion is that we should hold the past in a little more respect when we attempt to plumb its secrets. While the past entertains, ennobles, and expands quite readily, it enlightens only with delicate coaxing.

Looking for Wisdom

Although the past never repeats itself in detail, it is often viewed as having repetitive elements. People make the same kinds of decisions, face the same kinds of challenges, and suffer the same kinds of misfortune often enough for behavioral scientists to believe that they can detect recurrent patterns. Such faith prompts psychometricians to study the diagnostic secrets of ace clinicians, clinicians to look for correlates of aberrant behavior, brokers to hunt for harbingers of price increases, and dictators to ponder revolutionary situations. Their search usually has a logic paralleling that of multiple regression or correlation. A set of relevant cases is collected and each member is characterized on a variety of

From chapter 23 in *Judgment under Uncertainty: Heuristics and Biases*, ed. D. Kahneman, P. Slovic, and A. Tversky (New York: Cambridge University Press, 1982): 335–351. Reprinted with permission.

dimensions. The resulting matrix is scoured for significant relationships that might aid us in predicting the future. . . .

Formal Modeling

The *Daily Racing Form*, for example, offers the earnest handicapper some 100 pieces of information on each horse in any given race. The handicapper with a flair for data processing might commit to some computer's memory the contents of a bound volume of the *Form* and try to derive a formula predicting speed as a weighted sum of scores on various dimensions. For example:

$$\tilde{y} = b_1x_1 + b_2x_2 + b_3x_3 \tag{27.1}$$

where $\tilde{y}$ is our best guess at a horse's speed, x_1 is its percentage of victories in previous races, x_2 is its jockey's percentage of winning races, and x_3 is the weight it will carry in the present race. Assuming that standardized scores[1] are used, the weights (b_i) reflect the importance of the different factors. If $b_1 = 2b_2$, then a given change in the horse's percentage of wins affects our speed prediction twice as much as an equivalent change in the jockey's percentage of wins, because past performances have proved twice as sensitive to x_1 as x_2.

Sounds easy, but there are a thousand pitfalls. One emerges when the predictors (x_i) are correlated, as might (and in fact does) happen were winning horses to draw winning jockeys (or vice versa). In such cases of multicollinearity, each variable has some independent ability to explain past performance and the two have some shared ability. When the weights are determined, that shared explanatory capacity will somehow be split between the two. Typically, that split renders the (b_i) uninterpretable with any degree of precision. Thus the regression equation cannot be treated as a theory of horse racing, showing the importance of various factors.

A more modest theoretical goal would simply be to determine which factors are and which factors are not important, on the basis of how much each adds to our understanding of y. The logic here is that of stepwise regression; additional variables are added to the equation as long as they add something to its overall predictive (or explanatory) power. Yet even this minimalistic strategy can run afoul of multicollinearity. If many reflections of a particular factor (e.g., different aspects of breeding) are included, their shared explanatory ability may be divided up into such small pieces that no one aspect makes a "significant" contribution.

Of course, these nuances may be of relatively little interest to handicappers as long as the formula works well enough to help them somewhat in beating the odds. We scientist types, however, want wisdom as well as efficacy from our techniques. It is hard for us to give up interpreting weights. Regression procedures not only express, but also produce, understanding (or, at least, results) in a mechanical, repeatable fashion. Small wonder then that they have been pursued doggedly despite their limitations. One of the best documented pursuits has been in the study of clinical judgment. Clinical judgment is exercised by a radiologist who sorts X rays of ulcers into "benign" and "malignant," by a personnel officer who chooses the best applicants from a set of candidates, or by a crisis-center counselor who decides which callers threatening suicide are serious. In each of these examples, the diagnosis involves making a decision

on the basis of a set of cues or attributes. When, as in these examples, the decision is repetitive and all cases can be characterized by the same cues, it is possible to model the judge's decision-making policy statistically. One collects a set of cases for which the expert has made a summary judgment (e.g., benign, serious) and then derives a regression equation, like Equation 27.1, whose weights show the importance the judge has assigned to each cue.

Two decades of such policy-capturing studies persistently produced a disturbing pair of conclusions: (a) Simple linear models, using a weighted sum of the cues, did an excellent job of predicting judges' decisions, although (b) the judges claimed that they were using much more complicated strategies (L. R. Goldberg, 1968, 1970; Slovic & Lichtenstein, 1971). A commonly asserted form of complexity is called configural judgment, in which the diagnostic meaning of one cue depends upon the meaning of other cues (e.g., "that tone of voice makes me think 'not suicidal' unless the call comes in the early hours of the morning").

Two reasons for the conflict between measured and reported judgment policies have emerged from subsequent research, each with negative implications for the usefulness of regression modeling for "capturing" the wisdom of past decisions. One was the growing realization that combining enormous amounts of information in one's head, as required by such formulas, overwhelms the computational capacity of anyone but an idiot savant. A judge trying to implement a complex strategy simply would not be able to do so with great consistency. Indeed, it is difficult to learn and use even a non-configural, weighted-sum, decision rule when there are many cues or unusual relationships between the cues and predicted variable (Slovic, 1974).

The second realization that has emerged from clinical judgment research is that simple linear models are extraordinarily powerful predictors. A simple substantive theory indicating what variables people care about when making decisions may be all one needs to make pretty good predictions of their behavior. If some signs encourage a diagnosis or decision and others discourage it, simply counting the number of encouraging and discouraging signs will provide a pretty good guess at the individual's behavior. The result, however, will be a more modest theory than one can derive by flashy regression modeling (Fischhoff, Goitein, & Shapira, 1982). Thus, while the past seems to be right out there to be understood, our standard statistical procedures do not always tell us what we want to know. If not used carefully, they may mislead us, leaving us less wise than when we started. We are tempted to embrace highly complicated theories in their entirety, without realizing that their power comes from very simple underlying notions rather than from having captured the essence of the past.

Looking for Folly

Focus on Failure

Searching for wisdom in historic events requires an act of faith—a belief in the existence of recurrent patterns waiting to be discovered. Searching for wisdom in the behavior of historical characters requires a somewhat different act of

faith—confidence that our predecessors knew things we do not know. The first of these faiths is grounded in philosophy; it distinguishes those who view history as a social science, not an ideographic study of unique events. The second of these faiths is grounded in charity and modesty. It distinguishes those who hope to see further by standing on the shoulders of those who came before from those satisfied with standing on their faces. Aphorisms like "those who do not study the past are condemned to repeat it" suggest that faith in the wisdom of our predecessors is relatively rare.

An active search for folly is, of course, not without merit. Not only do individuals for whom things do not go right often have a lot of explaining to do, but such explanations are crucial to learning from their experience. By seeing how things went wrong, we hope to make them go right in the future. The quest for misfortunes to account for is hardly difficult. The eye, journalist, and historian are all drawn to disorder. An accident-free drive to the store or a reign without wars, depressions, or earthquakes is for them uneventful.

Although it has legitimate goals, focus on failure is likely to mislead us by creating a distorted view of the prevalence of misfortune. The perceived likelihood of events is determined in part by the ease with which they are imagined and remembered (Tversky & Kahneman, 1973, 11). Belaboring failures should, therefore, disproportionately enhance their perceived frequency in the past (and perhaps future).

It is also likely to promote an unbalanced appraisal of our predecessors' performance. The muckracker in each of us is drawn to stories of welfare cheaters or the "over-regulation" of particular environmental hazards (e.g., the Occupational Safety and Health Administration's infamous standard for a workplace toilet-seat design). We tend to forget, though, that any fallible, but not diabolical, decision-making system produces errors of both kinds. For every cheater garnering undeserved benefits, there are one or several or a fraction of cheatees, denied their rights by the same imperfect system. In fact, the two error rates are tied in a somewhat unintuitive fashion dependent upon the accuracy of judgment and the total resources available, that is, the percentage of eligible indigent or hazards that can be treated (Einhorn, 1978). Before rushing to criticize the welfare system for allowing a few cheaters, we should consider whether or not there might not be too few horror stories of that type, given the ratio of errors of commission to errors of omission.

In general, there is a good chance of being misled when we examine in isolation decisions that only "work out" on a percentage basis.

What Was the Problem?

There are other contexts in which errors in the small may look different when some larger context is considered. For example, we are taught that scientific theories should roll over dead once any inconsistent evidence is present. As a result, we are quick to condemn the folly of scientists who persist in their theories despite having been "proved" wrong. Kuhn (1962), however, argued that such local folly might be consistent with more global wisdom in the search for scientific knowledge. Others (e.g., Feyerabend, 1975; Lakatos, 1970) have, in fact, extolled the role of disciplined anarchy in the growth of understanding and have doubted the possibility of wisdom's emerging from orderly adher-

ence to any one favored research method. They argue that obstinate refusal to look at contrary evidence or to abandon apparently disconfirmed theories is often necessary to scientific progress.

The $125 million settlement levied against Ford Motor Company in the Pinto case made the company's decision to save a few dollars in the design of that car's fuel tank seem like folly. Yet in purely economic terms, a guaranteed saving of, say, $15 on each of 10 million Pintos makes the risk of a few large law suits seem like a more reasonable gamble. Since the judgment in this well-publicized suit was reduced to $6 million upon appeal, the company may actually be ahead in strict economic terms, despite having had worse come to worst. Where the company may be faulted is in seeing one larger context (the number of cars on which it would save money), but not another (the non-economic consequences of its decision). It seems not to have realized the impact that adverse publicity would have on Ford's image as a safety-conscious auto maker or on prices for used Pintos (although that price was borne by Pinto owners, not producers).

If reprobation is the name of the game, a mistake is a mistake. Yet, if one is interested in learning from the experience of others, it is important to determine what problem they were attempting to solve. Upon careful examination, many apparent errors prove to represent deft resolution of the wrong problem. For example, if it is to be criticized at all, Ford might be held guilty of tactical wisdom and strategic folly (or perhaps of putting institutional health over societal well-being).

This distinction is important, not only for evaluating the past, but also for knowing what corrective measures need to be taken in the future. Usually, tactical mistakes are easier to correct than strategic misunderstandings. Once we have properly characterized a situation, there may be a "book," recording conventional wisdom as accumulated through trial-and-error experience, or at least formulas for optimally combining the information at our disposal (Hexter, 1971). Baseball managers, for example, may either know that it has proven successful to have the batter sacrifice with a runner on first and no one out in a close game or else have the statistics needed to calculate how to "go with the percentages." These guides are, however, unhelpful or misleading if the real problem to be solved is maintaining morale (the runner has a chance to lead the league in stolen bases) or aiding the box office (the fans need to see some swinging). Studies of surprise attacks in international relations reveal that surprised nations have often done a good job of playing by their own book but have misidentified the arena in which they were playing (Ben Zvi, 1976; Lanir, 1978). In a sense, they were reading the wrong book; the better they read, the quicker they met their demise.

One reason for the difficulty posed by strategic problems is that they must be "thought through" analytically, without the benefit of cumulative (statistical) experience. A second limitation is that misconceptions are often widely shared within a decision-making group or community. One is consulted on decisions only after one has completed the catechism in the book. Recurrent pieces of advice for institutions interested in avoiding surprises are (a) set up several separate analytical bodies in order to provide multiple, independent looks at a problem or (b) appoint one member to serve as "devil's advocate" for

unpopular points of view (Janis, 1972). In practice, the first strategy may fail because shared misconceptions make the groups very like one another, creating redundancy rather than pluralism (Chan, 1979). The second fails because advocates either bow to group pressure or are ostracized if they take their unpopular positions seriously, even when those "extreme" positions do not drastically challenge group preconceptions.

Failure to distinguish between tactical and strategic decisions can also create an undeserved illusion of wisdom. Banks and insurance companies are usually considered to be extremely rational and adroit in their decision-making processes. Yet a closer look reveals that this reputation comes from their success in making highly repetitive, tactical decisions in which they almost cannot lose. Home mortgages and life insurance policies are issued on the basis of conservative interpretations of statistical tables acquired and adjusted through massive trial-and-error experience. These institutions' ventures into more speculative decisions requiring analytical, strategic decisions suggest that they are no smarter than the rest of us. Commercial banks lost large sums of money in the 1960s through unwise investments in real estate investment trusts; a similarly minute percentage of their overall decisions in the 1970s has chained the U.S. economy to the future of semisolvent Third World countries to whom enormous ($60+ billion) loans have been made. (Although this linkage may be for the long-range good of humanity, that was not necessarily the problem the banks were solving.) The slow and erratic response of insurance companies to changes in the economics of casualty insurance and their almost haphazard, non-analytical methods for dealing with many non-routine risks should leave the rest of us feeling not so stupid when compared with these vaunted institutions.

Hindsight: Thinking Backward?

If we know what has happened and what problem an individual was trying to solve, we should be in a position to exploit the wisdom of our own hindsight in explaining and evaluating his or her behavior. Upon closer examination, however, the advantages of knowing how things turned out may be oversold (Fischhoff, 1975). In hindsight, people consistently exaggerate what could have been anticipated in foresight. They not only tend to view what has happened as having been inevitable but also to view it as having appeared "relatively inevitable" before it happened. People believe that others should have been able to anticipate events much better than was actually the case. They even misremember their own predictions so as to exaggerate in hindsight what they knew in foresight (Fischhoff and Beyth, 1975).

As described by historian Georges Florovsky (1969):

> The tendency toward determinism is somehow implied in the method of retrospection itself. In retrospect, we seem to perceive the logic of the events which unfold themselves in a regular or linear fashion according to a recognizable pattern with an alleged inner necessity. So that we get the impression that it really could not have happened otherwise. (p. 369)

An apt name for this tendency to view reported outcomes as having been relatively inevitable might be *creeping determinism*, in contrast with philosophical determinism, the conscious belief that whatever happens has to happen.

One corollary tendency is to telescope the rate of historical processes, exaggerating the speed with which "inevitable" changes are consummated (Fischer, 1970). For example, people may be able to point to the moment when the latifundia were doomed, without realizing that they took two and a half centuries to disappear. Another tendency is to remember people as having been much more like their current selves than was actually the case (Yarrow, Campbell, & Burton, 1970). A third may be seen in Barraclough's (1972) critique of the historiography of the ideological roots of Nazism. Looking back from the Third Reich, one can trace its roots to the writings of many authors from whose writings one could not have projected Nazism. A fourth is to imagine that the participants in a historical situation were fully aware of its eventual importance ("Dear Diary, The Hundred Years' War started today," Fischer, 1970). A fifth is the myth of the critical experiment, unequivocally resolving the conflict between two theories or establishing the validity of one. In fact, "the crucial experiment is seen as crucial only decades later. Theories don't just give up, since a few anomalies are always allowed. Indeed, it is very difficult to defeat a research programme supported by talented and imaginative scientists" (Lakatos, 1970, pp. 157–158).

In the short run, failure to ignore outcome knowledge holds substantial benefits. It is quite flattering to believe, or lead others to believe, that we would have known all along what we could only know with outcome knowledge, that is, that we possess hindsightful foresight. In the long run, however, undetected creeping determinism can seriously impair our ability to judge the past or learn from it.

Consider decision makers who have been caught unprepared by some turn of events and who try to see where they went wrong by re-creating their pre-outcome knowledge state of mind. If, in retrospect, the event appears to have seemed relatively likely, they can do little more than berate themselves for not taking the action that their knowledge seems to have dictated. They might be said to add the insult of regret to the injury inflicted by the event itself. When second-guessed by a hindsightful observer, their misfortune appears as incompetence, folly, or worse.

In situations where information is limited and indeterminate, occasional surprises and resulting failures are inevitable. It is both unfair and self-defeating to castigate decision makers who have erred in fallible systems, without admitting to that fallibility and doing something to improve the system. According to historian Roberta Wohlstetter (1962), the lesson to be learned from American surprise at Pearl Harbor is that we must "accept the fact of uncertainty and learn to live with it. Since no magic will provide certainty, our plans must work without it" (p. 401).

When we attempt to understand past events, we implicitly test the hypotheses or rules we use both to interpret and to anticipate the world around us. If, in hindsight, we systematically underestimate the surprises that the past held and holds for us, we are subjecting those hypotheses to inordinately weak tests and, presumably, finding little reason to change them. Thus, the very outcome knowledge which gives us the feeling that we understand what the past was all about may prevent us from learning anything from it.

Protecting ourselves against this bias requires some understanding of the psychological processes involved in its creation. It appears that when we receive

outcome knowledge, we immediately make sense out of it by integrating it into what we already know about the subject. Having made this reinterpretation, the reported outcome now seems a more or less inevitable outgrowth of the reinterpreted situation. "Making sense" out of what we are told about the past is, in turn, so natural that we may be unaware that outcome knowledge has had any effect on us. Even if we are aware of there having been an effect, we may still be unaware of exactly what it was. In trying to reconstruct our foresightful state of mind, we will remain anchored in our hindsightful perspective, leaving the reported outcome too likely looking.

As a result, merely warning people about the dangers of hindsight bias has little effect (Fischhoff, 1977b). A more effective manipulation is to force oneself to argue against the inevitability of the reported outcomes, that is, try to convince oneself that it might have turned out otherwise. Questioning the validity of the reasons you have recruited to explain its inevitability might be a good place to start (Koriat, Lichtenstein, & Fischhoff, 1980; Slovic & Fischhoff, 1977). Since even this unusual step seems not entirely adequate, one might further try to track down some of the uncertainty surrounding past events in their original form. Are there transcripts of the information reaching the Pearl Harbor Command prior to 7 A.M. on December 7? Is there a notebook showing the stocks you considered before settling on Waltham Industries? Are there diaries capturing Chamberlain's view of Hitler in 1939? An interesting variant was Douglas Freeman's determination not to know about any subsequent events when working on any given period in his definitive biography of Robert E. Lee (Commager, 1965). Although admirable, this strategy does require some naive assumptions about the prevalence of knowledge regarding who surrendered at Appomattox.

Looking at All

Why Look?

Study of the past is predicated on the belief that if we look, we will be able to discern some interpretable patterns. Considerable research suggests that this belief is well founded. People seem to have a remarkable ability to find some order or meaning in even randomly produced data. One of the most familiar examples is the gamblers' fallacy. Our feeling is that in flipping a fair coin, four successive "heads" will be followed by a "tail" (Lindman & Edwards, 1961). Thus in our minds, even random processes are constrained to have orderly internal properties. Kahneman and Tversky (1972, 3) have suggested that of the 32 possible sequences of six binary events only 1 actually looks "random."

Although the gamblers' fallacy is usually cited in the context of piquant but trivial examples, it can also be found in more serious attempts to explain historical events. For example, after cleverly showing that Supreme Court vacancies appear more or less at random (according to a Poisson process), with the probability of at least one vacancy in any given year being .39, Morrison (1977) claimed that:

> [President] Roosevelt announced his plan to pack the Court in February, 1937, shortly after the start of his fifth year in the White House. 1937 was

> also the year in which he made his first appointment to the Court. That he had this opportunity in 1937 should come as no surprise, because the probability that he would go five consecutive years without appointing one or more justices was but .08, or one chance in twelve. In other words, when Roosevelt decided to change the Court by creating additional seats, the odds were already *eleven to one in his favor* that he would be able to name one or more justices by traditional means that very year. (pp. 143–144)

However, if vacancies do appear at random, then this reasoning is wrong. It assumes that the probabilistic process creating vacancies, like that governing coin flips, has a memory and a sense of justice, as if it knows that it is moving into the fifth year of the Roosevelt presidency and that it "owes" FDR a vacancy. However, on January 1, 1937, the past four years were history, and the probability of at least one vacancy in the coming year was still .39 (Fischhoff, 1978).

Feller (1968) offers the following anecdote involving even higher stakes: Londoners during the blitz devoted considerable effort to interpreting the pattern of German bombing, developing elaborate theories of where the Germans were aiming (and when to take cover). However, when London was divided up into small, contiguous geographic areas, the frequency distribution of bomb-hits per area was almost a perfect approximation of the Poisson distribution. Kates (1962) suggests that natural disasters constitute another category of consequential events where (threatened) laypeople see order when experts see randomness.

One secret to maintaining such beliefs is failure to keep complete enough records to force ourselves to confront irregularities. Historians acknowledge the role of missing evidence in facilitating their explanations with comments like "the history of the Victorian Age will never be written. We know too much about it. For ignorance is the first requisite of the historian—ignorance which simplifies and clarifies, which selects and omits, with placid perfection unattainable by the highest art" (Strachey, 1918, preface).

Even where records are available and unavoidable, we seem to have a remarkable ability to explain or provide a causal interpretation for whatever we see. When events are produced by probabilistic processes with intuitive properties, random variation may not even occur to us as a potential hypothesis. For example, the fact that athletes chastised for poor performance tend to do better the next time out fits our naive theories of reward and punishment. This handy explanation blinds us to the possibility that the improvement is due instead to regression to those players' mean performance (Furby, 1973; Kahneman & Tversky, 1973, 4).

Fama (1965) has forcefully argued that the fluctuations of stock-market prices are best understood as reflecting a random walk process. Random walks, however, have even more unintuitive properties than the binary processes to which they are formally related (Carlsson, 1972). As a result, we find that market analysts have an explanation for every change in price, whether purposeful or not. Some explanations, like those shown in figure 27.1, are inconsistent;[2] others seem to deny the possibility of any random component, for example, that ultimate fudge factor, the "technical adjustment."

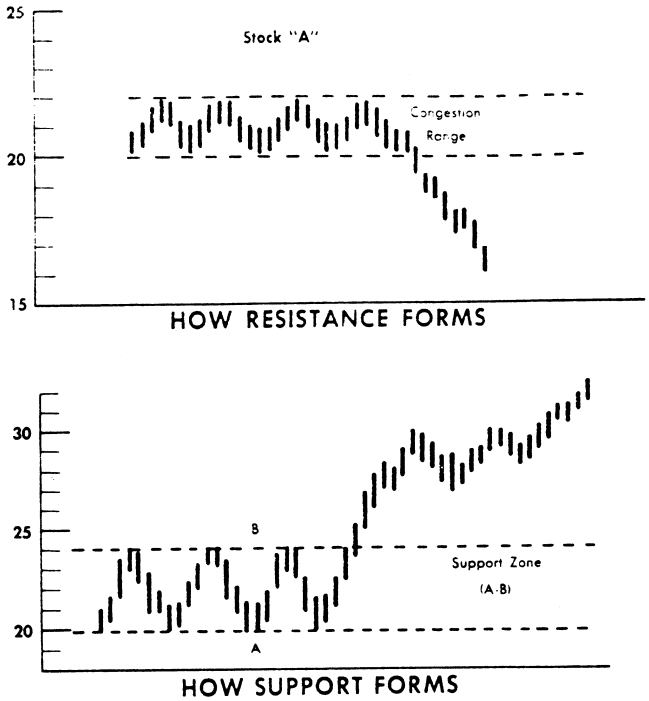


Figure 27.1
Two examples of cues used to identifying precursors of past shifts in stock prices: formation of resistance and formation of support. However, one might argue that prior to the dramatic shifts at their respective ends, these two patterns were essentially identical. In this light, an undulating pattern neither predicts nor explains anything in these data.

The pseudopower of our explanations can be illustrated by analogy with regression analysis. Given a set of events and a sufficiently large or rich set of possible explanatory factors, one can always derive postdictions or explanations to any desired degree of tightness. In regression terms, by expanding the set of independent variables one can always find a set of predictors with any desired correlation with the independent variable. The price one pays for overfitting is, of course, shrinkage, failure of the derived rule to work on a new sample of cases. The frequency and vehemence of methodological warnings against overfitting suggest that correlational overkill is a bias that is quite resistant to even extended professional training (for references, see Fischhoff & Slovic, 1980).

An overfitted theory is like a suit tailored so precisely to one individual in one particular pose that it will not fit anyone else or even that same individual in the future or even in the present if new evidence about him comes to light (e.g., if he lets out his breath to reveal a potbelly). A historian who had built an airtight case accounting for all available evidence in explaining how the Bolsheviks won might be in a sad position were the USSR to release suppressed documents showing that the Mensheviks were more serious adversaries than had previously been thought. The price that investment analysts pay for over-

fitting is their long-run failure to predict any better than market averages (Dreman, 1979)—although the cynic might say that they actually make their living through the generation of hope (and commissions).[3]

Overfitting occurs because of capitalization on chance fluctuations. If measurement is sufficiently fine, two cases differing on one variable will also differ on almost any other variable one chooses to name. As a result, one can calculate a non-zero (actually, in this case, perfect) correlation between the two variables and derive an "interesting" substantive theory. Processes analogous to this two-dimensional case work with any m observations in the n-space defined by our set of possible explanatory concepts.

In these examples, the data are fixed and undeniable, while the set of possible explanations is relatively unbounded; one hunts until one finds an explanation that fits. Another popular form of capitalization on chance leaves the set of explanations fixed (usually at one candidate) and sifts through data until supporting evidence is found. Although the crasser forms of this procedure are well known, others are more subtle and even somewhat ambiguous in their characterization. For example, you run an experiment and fail to receive an anticipated result. Thinking about it, you note an element of your procedure that might have mitigated the effect of the manipulated variable. You correct that; again no result but, again, a possible problem. Finally, you (or your subjects) get it right and the anticipated effect is obtained. Now, is it right to perform your statistical test on that nth sample (for which it shows significance) or the whole lot of them? Had you done the right experiment first, the question would not even have arisen. Or, as a toxicologist, you are "certain" that exposure to chemical X is bad for one's health, so you compare workers who do and do not work with it in a particular plant for bladder cancer, but still no effect. So you try intestinal cancer, emphysema, dizziness, and so on, until you finally get a significant difference in skin cancer. Is that difference meaningful? Of course the way to test these explanations or theories is by replication on new samples. That step, unfortunately, is seldom taken and often is not possible for technical or ethical reasons (Tukey, 1977).

Related complications can arise even with fixed theories and data sets. Diaconis (1978) notes the difficulty of evaluating the amount of surprise in ESP results, even in the rare cases in which they have been obtained in moderately supervised settings, because the definition of the sought event keeps shifting. "A major key to B.D.'s success was that he did not specify in advance the result to be considered surprising. The odds against a coincidence *of some sort* are dramatically less than those against any prespecified *particular one* of them" (p. 132).[4]

Tufte and Sun (1975) discovered that the existence or non-existence of bellwether precincts depends upon the creativity and flexibility allowed in defining the event (for what office? in what elections? how good is good? are precincts that miss consistently to be included?). They are commonly believed to exist because we have an uncommonly good ability to find a signal even in total noise.

Have We Seen Enough?

Given that we are almost assured of finding something interpretable when we look at the past, our next question becomes, "Have we understood it?" The

hindsight research described earlier suggests that we are not only quick to find order but also poised to feel that we knew it all along in some way or would have been able to predict the result had we been asked in time. Indeed, the ease with which we discount the informativeness of anything we are told makes it surprising that we ever ask the past, or any other source, many questions. This tendency is aggravated by tendencies (a) not to realize how little we know or are told, leaving us unaware of what questions we should be asking in search of surprising answers (Fischhoff, Slovic, & Lichtenstein, 1977, 1978) and (b) to draw far-reaching conclusions from even small amounts of unreliable data (Kahneman & Tversky, 1973, 4; Tversky & Kahneman, 1971, 2).

Any propensity to look no further is encouraged by the norm of reporting history as a good story, with all the relevant details neatly accounted for and the uncertainty surrounding the event prior to its consummation summarily buried, along with any confusion the author may have felt (Gallie, 1964; Nowell-Smith, 1970). Just one of the secrets to doing this is revealed by Tawney (1961): "Historians give an appearance of inevitability to an existing order by dragging into prominence the forces which have triumphed and thrusting into the background those which they have swallowed up" (p. 177).[5]

Although an intuitively appealing goal, the construction of coherent narratives exposes the reader to some interesting biases. A completed narrative consists of a series of somewhat independent links, each fairly well established. The truth of the narrative depends upon the truth of the links. Generally, the more links there are and the more detail there is in each link, the less likely the story is to be correct in its entirety. However, Slovic, Fischhoff, and Lichtenstein (1976) have found that adding detail to an event description can increase its perceived probability of occurrence, evidently by increasing its thematic unity. Bar-Hillel (1973) found that people consistently exaggerate the probability of the conjunction of a series of likely events. For example, her subjects generally preferred a situation in which they would receive a prize if seven independent events each with a probability of .90 were to occur to a situation in which they would get the same prize if a fair coin fell on "heads." The probability of the compound event is less than .50, whereas the probability of the single event is .50. In other words, uncertainty seems to accumulate at much too slow a rate.

What happens if the sequence includes one or a few weak or unlikely links? The probability of its weakest link should set an upper limit on the probability of an entire narrative. Coherent judgments, however, may be compensatory, with the coherence of strong links "evening out" the incoherence of weak links. This effect is exploited by attorneys who bury the weakest link in their arguments near the beginning of their summations and finish with a flurry of convincing, uncontestable arguments.

Coles (1973) presents a delicious example of the overall coherence of a story obscuring the unlikelihood of its links: Freud's most serious attempt at psychohistory was his biography of Leonardo da Vinci. For years, Freud had sought the secret to understanding Leonardo, whose childhood and youth were basically unknown. Finally, he discovered a reference by Leonardo to a recurrent memory of a vulture touching his lips while he was in the cradle. Noting the identity of the Egyptian hieroglyphs for "vulture" and "mother" and other

circumstantial evidence, Freud went on to build an imposing and coherent analysis of Leonardo. While compiling the definitive edition of Freud's works, however, the editor discovered that the German translation of Leonardo's recollection (originally in Italian) that Freud had used was in error, and that it was a kite not a vulture that had stroked his lips. Despite having the key to Freud's analysis destroyed, the editors decided that the remaining edifice was strong enough to stand alone. As Hexter (1971) observed, "Partly because writing bad history is pretty easy, writing very good history is rare" (p. 59).

Conclusion

What general lessons can we learn about the study of the past, beyond the fact that understanding is more elusive than may often be acknowledged?

Presentism

Inevitably, we are all captives of our present personal perspective. We know things that those living in the past did not. We use analytical categories (e.g., feudalism, Hundred Years War) that are meaningful only in retrospect (E. A. R. Brown, 1974). We have our own points to prove when interpreting a past that is never sufficiently unambiguous to avoid the imposition of our ideological perspective (Degler, 1976). Historians do "play new tricks on the dead in every generation" (Becker, 1935).

There is no proven antidote to presentism. Some partial remedies can be generalized from the discussion of how to avoid hindsight bias when second-guessing the past. Others appear in almost any text devoted to the training of historians. Perhaps the most general messages seem to be (a) knowing ourselves and the present as well as possible; "the historian who is most conscious of his own situation is also most capable of transcending it" (Benedetto Croce, quoted in Carr, 1961, p. 44); and (b) being as charitable as possible to our predecessors; "the historian is not a judge, still less a hanging judge" (Knowles, quoted in Marwick, 1970, p. 101).

Methodism

In addition to the inescapable prison of our own time, we often further restrict our own perspective by voluntarily adopting the blinders that accompany strict adherence to a single scientific method. Even when used judiciously, no one method is adequate for answering many of the questions we put to the past. Each tells us something and misleads us somewhat. When we do not know how to get the right answer to a question, an alternative epistemology is needed: Use as broad a range of techniques or perspectives as possible, each of which enables us to avoid certain kinds of mistakes. This means a sort of interdisciplinary cooperation and respect different from that encountered in most attempts to commingle two approaches. Matches or mismatches like psychohistory too often are attempted by advocates insensitive to the pitfalls in their adopted fields (Fischhoff, 1981). Hexter (1971) describes the historians involved in some such adventures as "rats jumping aboard intellectually sinking ships" (p. 110).

Learning

Returning to Benson (1972), if we want the past to serve the future, we cannot treat it in isolation. The rules we use to explain the past must also be those we use to predict the future. We must cumulate our experience with a careful eye to all relevant tests of our hypotheses. One aspect of doing this is compiling records that can be subjected to systematic statistical analysis: A second is keeping track of the deliberations preceding our own decisions, realizing that the present will soon be past and that a well-preserved record is the best remedy to hindsight bias: A third is making predictions that can be evaluated; one disturbing lesson from the Three Mile Island nuclear accident is that it is not entirely clear what that ostensibly diagnostic event told us about the validity of the Reactor Safety Study (U.S. Nuclear Regulatory Commission, 1975) that attempted to assess the risks from nuclear power: A fourth aspect is getting a better idea of the validity of our own feelings of confidence, insofar as confidence in present knowledge controls our pursuit of new information and interpretations (Fischhoff, Slovic, & Lichtenstein, 1977). Thus, we should structure our lives so as to facilitate learning.

Indeterminacy

In the end, though, there may be no answers to many of the questions we are posing. Some are ill-formed. Others just cannot be answered with existing or possible tools. As much as we would like to know "how the pros do it," there may be no way statistically to model experts' judgmental policies to the desired degree of precision with realistic stimuli. Our theories are often of "such complexity that no single quantitative work could even begin to test their validity" (O'Leary et al., 1974, p. 228). When groups we wish to compare on one variable also differ on another, there is no logically sound procedure for equating them on that nuisance variable (Meehl, 1970). When we have tried many possible explanations on a fixed set of data, there is no iron-clad way of knowing just how many degrees of freedom we have used up, just how far we have capitalized on chance (Campbell, 1975). When we use multiple approaches, the knowledge they produce never converges neatly. In the end, we may have to adopt Trevelyan's philosophical perspective that "several imperfect readings of history are better than none at all" (cited in Marwick, 1970, p. 57).

Notes

This is a revised version of the paper "For Those Condemned to Study the Past: Reflections on Historical Judgment," in R. A. Shweder and D. W. Fiske (Eds.), *New Directions for Methodology of Behavioral Science: Fallible Judgment in Behavioral Research*. San Francisco: Jossey-Bass, 1980. Reprinted by permission.

1. To standardize scores on a particular variable, one subtracts the mean of all scores from each score and then divides by the standard deviation. The result is a set of scores with a mean of 0 and a standard deviation of 1.
2. One of my favorite contrasts is that when the market rises following good economic news, it is said to be responding to the news; if it falls, that is explained by saying that the good news had already been discounted.
3. A friend once took a course in reading form charts from a local brokerage. Each session involved the teaching of 10–12 new cues. When the course ended, five sessions and 57 cues later, the instructor was far from exhausting his supply.

4. Diaconis continues, "To further complicate any analysis, several such ill-defined experiments were often conducted simultaneously, inter-acting with one another. The young performer electrified his audience. His frequently completely missed guesses were generally regarded with sympathy, rather than doubt; and for most observers they seemed only to confirm the reality of B.D.'s unusual powers."
5. Such strategies may affect the spirit as well as the mind, by subjectively enhancing the strength and stability of the status quo and reducing its apparent capacity for change (Marković, 1970).

References

Bar-Hillel, M. (1973). On the subjective probability of compound events. *Organizational Behavior and Human Performance, 9*, 396–406.

Barraclough, G. (1972). Mandarins and Nazis. *New York Review of Books, 19*, 37–42.

Becker, C. (1935). Everyman his own historian. *American Historical Review, 40*, 221–236.

Ben Zvi, A. (1976). Hindsight and foresight: A conceptual framework for the analysis of surprise attacks. *World Politics, 28*, 381–395.

Benson, L. (1972). *Toward the scientific study of history: Selected essays*. Philadelphia: Lippincott.

Brown, E. A. R. (1974). The tyranny of a construct: Feudalism and historians of medieval Europe. *American Historical Review, 79*, 1063–1088.

Campbell, D. T. (1975). Degrees of freedom: And the case study. *Comparative Political Studies, 8*, 178–193.

Carr, E. H. (1961). *What is history?* London: Macmillan.

Carlsson, G. (1972). Random walk effects in behavioral data. *Behavioral Science, 17*, 430–437.

Chan, S. (1979). The intelligence of stupidity: Understanding failures in strategic warning. *American Political Science Review, 73*, 171–180.

Coles, R. (1973). Shrinking history. *New York Review of Books, Part I*, February 22, 1973, pp. 20, 15–21; *Part II*, March 8, 1973, pp. 20, 25–29.

Commager, H. S. (1965). *The nature and study of history*. Columbus, Ohio: Merrill.

Degler, C. N. (1976). Why historians change their minds. *Pacific Historical Review, 48*, 167–189.

Diaconis, P. (1978). Statistical problems in ESP research. *Science, 201*, 131–136.

Dreman, D. (1979). *Contrarian investment strategy*. New York: Random House.

Fama, E. F. (1965). Random walks in stock market prices. *Financial Analysts Journal, 21*, 55–60.

Feller, W. (1968). *An introduction to probability theory and its applications* (3d Ed., Vol. 1). New York: Wiley.

Feyerabend, P. (1975). *Against method*. New York: NLB (Schocken).

Fischer, D. H. (1970). *Historians' fallacies*. New York: Harper and Row.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288–299.

Fischhoff, B. (1981). No man is a discipline. In J. Harvey (Ed.), *Cognition, social behavior and the* environment. Hillsdale, NJ: Erlbaum.

Fischhoff, B. (1982). Latitude and platitudes. How much credit do people deserve? In G. Ungson and D. Braunstein (Eds.), *New directions in decision making*. New York: Kent.

Fischhoff, B., & Beyth, R. (1975). "I knew it would happen"—Remembered probabilities of once-future things. *Organizational Behavior and Human Performance, 13*, 1–16.

Fischhoff, B., Goitein, B., & Shapira, Z. (1982). The expected utility of expected utility approaches. In N. T. Feather (Ed.), Expectations and actions: Expectancy-value models in psychology (pp. 315–339). Hillsdale, NJ: Erlbaum.

Fischhoff, B., & Slovic, P. (1980). A little learning. . . : Confidence in multicue judgment. In R. Nickerson (Ed.), *Attention and Performance VIII*. Hillsdale, NJ: Erlbaum.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 552–564.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 330–334.

Florovosky, G. (1969). The study of the past. In R. H. Nash (Ed.), *Ideas of history* (Vol. 2). New York: Dutton.

Furby, L. (1973). Interpreting regression toward the mean in developmental research. *Developmental Psychology, 8*, 172–179.

Gallie, W. B. (1964). *Philosophy and the historical understanding*. London: Chatto & Windus.

Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgements. *American Psychologist, 23*, 483–496.

Goldberg, L. R. (1970). Man vs. model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin, 73*, 422–432.

Hexter, J. H. (1971). *The history primer*. New York: Basic Books.

Janis, I. (1972). *Victims of groupthink*. Boston: Houghton Mifflin.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.

Kates, R. W. (1962). *Hazard and choice perception in flood plain management* (Research Paper No. 78). Chicago: University of Chicago, Department of Geography.

Koriat, A., Lichtenstein, S., & Fischoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107–118.

Kuhn, T. (1962). *The structure of scientific revolution*. Chicago: University of Chicago Press.

Lakatos, I. (1970). Falsification and scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of scientific knowledge*. Cambridge: Cambridge University Press.

Lanir, Z. (1978). *Critical reevaluations of the strategic intelligence methodology*. Tel Aviv: Center for Strategic Studies, Tel Aviv University.

Lindman, H. G., & Edwards, W. (1961). Supplementary report: Learning the gambler's fallacy. *Journal of Experimental Psychology, 37*, 2098–2110.

Marwick, A. (1970). *The nature of history*. London: Macmillan.

Meehl, P. E. (1970). Nuisance variables and the ex post facto design. In M. Radner & S. Winokur (Eds.), *Minnesota studies in the philosophy of science*. Minneapolis: University of Minnesota Press.

Morrison, R. J. (1977). Franklin D. Roosevelt and the Supreme Court: An example of the use of probability theory in political history. *History and Theory, 16*, 137–146.

Nowell-Smith, P. H. (1970). Historical explanation. In H. E. Kiefer & M. K. Munitz (Eds.), *Mind, science and history*. Albany, NY: State University of New York Press.

O'Leary, M. K., Coplin, W. D., Shapiro, H. B., & Dean, D. (1974). The quest for relevance. *International Studies Quarterly, 18*, 211–237.

Slovic, P. (1974). Hypothesis testing in the learning of positive and negative linear functions. *Organizational Behavior and Human Performance, 11*, 368–376.

Slovic, P., & Fischoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 544–551.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6*, 649–744.

Slovic, P., Fischoff, B., & Lichtenstein, S. (1976). Cognitive processes and societal risk taking. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior*. Hillsdale, NJ: Erlbaum.

Strachey, L. (1918). *Eminent Victorians*. New York: Putnam.

Tawney, R. H. (1961). *The agrarian problems in the sixteenth century*. New York: Franklin.

Tufte, E. R., & Sun, R. A. (1975). Are there bellwether electoral districts? *Public Opinion Quarterly, 39*, 1–18.

Tukey, J. W. (1977). Some thoughts on clinical trials, especially the problems of multiplicity. *Science, 198*, 679–690.

Tversky, A., & Kahneman, D. (1973). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Tversky, A., & Kahneman, D. (1971). The belief in the "law of small numbers." *Psychological Bulletin, 76*, 105–110.

U.S. Nuclear Regulatory Commision (1975). *Reactor safety study: An assessment of accident risks in U.S. commercial nuclear power plants* (WASH 1400 [NUREG-75/014]). Washington, DC: NRC.

Wohlstetter, R. (1962). *Pearl Harbor: Warning and decision*. Stanford, CA: Stanford University Press.

Yarrow, M., Campbell, J. D., & Burton, R. V. (1970). Recollections of childhood: A study of the retrospective method. *Monographs of the Society for Research in Child Development, 35*.

PART XIII

Evolutionary Approaches

Chapter 28

Adaptations, Exaptations, and Spandrels

David M. Buss, Martie G. Haselton, Todd K. Shackelford, April L. Bleske, and Jerome C. Wakefield

---

Over the past decade, evolutionary psychology has emerged as a prominent new theoretical perspective within the field of psychology. Evolutionary psychology seeks to synthesize the guiding principles of modern evolutionary theory with current formulations of psychological phenomena (Buss, 1995; Daly & Wilson, 1988; Pinker, l997b; Symons, 1987; Tooby & Cosmides, 1992). The concepts of adaptation and natural selection are central to evolutionary approaches and, therefore, have figured prominently in this emerging perspective. At the same time, criticisms have been leveled at the concept of adaptation and the importance of natural selection, especially as they are applied to human behavior. In particular, Gould (1991), in an influential and widely cited analysis, suggested that "exaptation," a feature not arising as an adaptation for its current function but rather co-opted for new purposes, may be a more important concept for the emerging paradigm of evolutionary psychology.

Psychologists in cognitive, developmental, social, personality, and clinical psychology are increasingly incorporating the evolutionary concepts of adaptation and exaptation in their theoretical frameworks and empirical research (e.g., Buss, 1994; Cosmides, 1989; Cosmides & Tooby, 1994; Daly & Wilson, 1988; Kenrick & Keefe, 1992; Lilienfeld & Marino, 1995; MacNeilage, 1997; Piattelli-Palmarini, 1989; Pinker & Bloom, 1992; Richters & Cicchetti, 1993; Sedikedes & Skowronski, 1997; Wakefield, 1992, 1999). Much confusion exists, however, about what these central concepts mean, how they should be distinguished, and how they are to be applied to psychological phenomena.

The confusion can be traced to several factors. First, psychologists typically receive no formal training in evolutionary biology and, therefore, cannot be expected to wade through what has become a highly technical field. Second, although evolutionary theorizing about humans has a long history (e.g., Baldwin, 1894; Darwin, 1859/1958; James, 1890/1962; Jennings, 1930; Morgan, 1896; Romanes, 1889), the empirical examination within psychology of evolutionary hypotheses regarding human psychological mechanisms is much more recent, and confusion often inheres in newly emerging approaches as practitioners struggle, often with many false starts, to use an incipient set of theoretical tools.[1] Third, psychologists dating back to Darwin's time have had a history of wariness about evolutionary approaches and, therefore, often have avoided a serious consideration of their potential utility. Fourth, there are genuine

From *American Psychologist* 53, no. 5 (May 1998): 533–548. Reprinted with permission.

differences in scientific opinion about which concepts should be used, what the concepts actually mean, and how they should be applied. This article seeks to provide psychologists with a guide to the basic concepts involved in the current dispute over evolutionary explanations and to clarify the role that each of these concepts plays in an evolutionary approach to human psychology.

The Evolutionary Process

The process of evolution—changes over time in organic structure—was hypothesized to occur long before Charles Darwin (1859/1958) formulated his theory of evolution. What the field of biology lacked, however, was a causal mechanism to account for these changes. Darwin supplied this causal mechanism in the form of natural selection.

Darwin's task was more difficult than it might appear at first. He wanted not only to explain why life-forms have the characteristics they do and why these characteristics change over time but also to account for the particular ways in which they change. He wanted to explain how new species emerge (hence the title of his book, *On the Origin of Species by Means of Natural Selection*; Darwin, 1859/1958) as well as how others vanish. Darwin wanted to explain why the component parts of animals—the long necks of giraffes, the wings of birds, the trunks of elephants, and the proportionately large brains of humans—exist in the particular forms they do. In addition, he wanted to explain the apparent purposive quality of these complex organic forms, or why they seem to function to help organisms to accomplish specific tasks.

Darwin's (1859/1958) answer to all these puzzles of life was the theory of natural selection. Darwin's theory of natural selection had three essential ingredients: variation, inheritance, and selection. Animals within a species vary in all sorts of ways, such as wing length, trunk strength, bone mass, cell structure, fighting ability, defensive maneuverability, and social cunning. This variation is essential for the process of evolution to operate. It provides the raw materials for evolution.

Only some of these variations, however, are reliably passed down from parents to offspring through successive generations. Other variations, such as a wing deformity caused by a chance environmental accident, are not inherited by offspring. Only those variations that are inherited play a role in the evolutionary process.

The third critical ingredient of Darwin's (1859/1958) theory was selection. Organisms with particular heritable attributes produce more offspring, on average, than those lacking these attributes because these attributes help to solve specific problems and thereby contribute to reproduction in a particular environment. For example, in an environment in which the primary food source is nut-bearing trees or bushes, some finches with a particular shape of beak might be better able to crack nuts and get at their meat than finches with alternative beak shapes. More finches that have the beaks better shaped for nut-cracking survive than those with beaks poorly shaped for nut-cracking. Hence, those finches with more suitably shaped beaks are more likely, on average, to live long enough to pass on their genes to the next generation.

Organisms can survive for many years, however, and still fail to contribute inherited qualities to future generations. To pass on their qualities, they must reproduce. Differential reproductive success, by virtue of the possession of heritable variants, is the causal engine of evolution by natural selection. Because survival is usually necessary for reproduction, survival took on a critical role in Darwin's (1859/1958) theory of natural selection.

Darwin (1859/1958) envisioned two classes of evolved variants—one playing a role in survival and one playing a role in reproductive competition. For example, among humans, sweat glands help to maintain a constant body temperature and thus presumably help humans to survive. Humans' tastes for sugar and fat presumably helped to guide their ancestors to eat certain foods and to avoid others and thus helped them to survive. Other inherited attributes aid more directly in reproductive competition and are said to be sexually selected (Darwin, 1871/1981). The elaborate songs and brilliant plumage of various bird species, for example, help to attract mates, and hence to reproduce, but may do nothing to enhance the individual's survival. In fact, these characteristics may be detrimental to survival by carrying large metabolic costs or by alerting predators.

In summary, although differential reproductive success of inherited variants was the crux of Darwin's (1859/1958) theory of natural selection, he conceived of two classes of variants that might evolve—those that help organisms survive (and thus indirectly help them to reproduce) and those that more directly help organisms in reproductive competition. The theory of natural selection unified all living creatures, from single-celled amoebas to multicellular mammals, into one grand tree of descent. It also provided for the first time a scientific theory to account for the exquisite design and functional nature of the component parts of each of these species.

In its modern formulation, the evolutionary process of natural selection has been refined in the form of inclusive fitness theory (Hamilton, 1964). Hamilton reasoned that classical fitness—a measure of an individual's direct reproductive success in passing on genes through the production of offspring—was too narrow to describe the process of evolution by selection. He proposed that a characteristic will be naturally selected if it causes an organism's genes to be passed on, regardless of whether the organism directly produces offspring. If a person helps a brother, a sister, or a niece to reproduce and nurture offspring, for example, by sharing resources, offering protection, or helping in times of need, then that person contributes to the reproductive success of his or her own genes because kin tend to share genes and, moreover, contributes to the reproductive success of genes specifically for brotherly, sisterly, or niecely assistance (assuming that such helping is partly heritable and, therefore, such genes are likely to be shared by kin). The implication of this analysis is that parental care—investing in one's own children—is merely a special case of caring for kin who carry copies of one's genes in their bodies. Thus, the notion of classical fitness was expanded to inclusive fitness.

Technically, inclusive fitness is not a property of an individual organism but rather a property of its actions or effects (Hamilton, 1964; see also Dawkins, 1982). Inclusive fitness can be calculated from an individual's own reproductive

success (classical fitness) plus the effects the individual's actions have on the reproductive success of his or her genetic relatives, weighted by the appropriate degree of genetic relatedness.

It is critical to keep in mind that evolution by natural selection is not forward looking or intentional. A giraffe does not notice juicy leaves stirring high in a tree and "evolve" a longer neck. Rather, those giraffes that happen to have slightly longer necks than other giraffes have a slight advantage in getting to those leaves. Hence, they survive better and are more likely to live to pass on genes for slightly longer necks to offspring. Natural selection acts only on those variants that happen to exist. Evolution is not intentional and cannot look into the future to foresee distant needs.

Products of the Evolutionary Process: Adaptations, By-Products, and Random Effects

In each generation, the process of selection acts like a sieve (Dawkins, 1996). Variants that interfere with successful solutions to adaptive problems are filtered out. Variants that contribute to the successful solution of an adaptive problem pass through the selective sieve. Iterated over thousands of generations, this filtering process tends to produce and maintain characteristics that interact with the physical, social, or internal environment in ways that promote the reproduction of individuals who possess the characteristics or the reproduction of the individuals' genetic relatives (Dawkins, 1982; Hamilton, 1964; Tooby & Cosmides, 1990a; Williams, 1966). These characteristics are called adaptations.

There has been much debate about the precise meaning of adaptation, but we offer a provisional working definition. An *adaptation* may be defined as an inherited and reliably developing characteristic that came into existence as a feature of a species through natural selection because it helped to directly or indirectly facilitate reproduction during the period of its evolution (after Tooby & Cosmides, 1992). Solving an adaptive problem—that is, the manner in which a feature contributes to reproduction—is the function of the adaptation. There must be genes for an adaptation because such genes are required for the passage of the adaptation from parents to offspring. Adaptations, therefore, are by definition inherited, although environmental events may play a critical role in their ontogenetic development.

Ontogenetic events play a profound role in several ways. First, interactions with features of the environment during ontogeny (e.g., certain placental nutrients, aspects of parental care) are critical for the reliable development and emergence of most adaptations. Second, input during development may be required to activate existing mechanisms. There is some evidence, for example, that experience in committed sexual relationships activates sex-linked jealousy adaptations (Buss, Larsen, Westen, & Semmelroth, 1992). Third, developmental events may channel individuals into one of several alternative adaptive paths specified by evolved decision rules. Lack of an investing father during the first several years of life, for example, may incline individuals toward a short-term mating strategy, whereas the presence of an investing father may shift individuals toward a long-term mating strategy (e.g., Belsky, Steinberg, & Draper, 1991; for alternative theories, see Buss & Schmitt, 1993; Gangestad & Simpson,

1990). Fourth, environmental events may disrupt the emergence of an adaptation in a particular individual, and thus the genes for the adaptation do not invariantly result in its intact phenotypic manifestation. Fifth, the environment during development may affect where in the selected range someone falls, such as which language a person speaks or how anxious a person tends to be. Developmental context, in short, plays a critical role in the emergence and activation of adaptations (see DeKay & Buss, 1992, for a more extended discussion of the role of context).

To qualify as an adaptation, however, the characteristic must reliably emerge in reasonably intact form at the appropriate time during an organism's life. Furthermore, adaptations tend to be typical of most or all members of a species, with some important exceptions, such as characteristics that are sex-linked, that exist only in a subset because of frequency-dependent selection, or that exist because of temporally or spatially varying selection pressures.

Adaptations need not be present at birth. Many adaptations develop long after birth. Bipedal locomotion is a reliably developing characteristic of humans, but most humans do not begin to walk until a year after birth. The breasts of women and a variety of other secondary sex characteristics reliably develop, but they do not start to develop until puberty.

The characteristics that make it through the filtering process in each generation generally do so because they contribute to the successful solution of adaptive problems—solutions that either are necessary for reproduction or enhance relative reproductive success. Solutions to adaptive problems can be direct, such as a fear of dangerous snakes that solves a survival problem or a desire to mate with particular members of one's species that helps to solve a reproductive problem. They can be indirect, as in a desire to ascend a social hierarchy, which many years later might give an individual better access to mates. Or they can be even more indirect, such as when a person helps a brother or a sister, which eventually helps that sibling to reproduce or nurture offspring. Adaptive solutions need not invariably solve adaptive problems in order to evolve. The human propensity to fear snakes, for example, does not inevitably prevent snakebites, as evidenced by the hundreds of people who die every year from snakebites (Than-Than et al., 1988). Rather, adaptive designs must provide reproductive benefits on average, relative to their costs and relative to alternative designs available to selection, during the period of their evolution.

Each adaptation has its own period of evolution. Initially, a mutation occurs in a single individual. Most mutations disrupt the existing design of the organism and hence hinder reproduction. If the mutation is helpful to reproduction, however, it will be passed down to the next generation in greater numbers. In the next generation, therefore, more individuals will possess the characteristic. Over many generations, if it continues to be successful, the characteristic will spread among the population. In sum, natural selection is the central explanatory concept of evolutionary theory, and adaptation refers to any functional characteristic whose origin or maintenance must be explained by the process of natural selection.[2]

Most adaptations, of course, are not caused by single genes. The human eye, for example, takes thousands of genes to construct. An adaptation's environment of evolutionary adaptedness (EEA) refers to the cumulative selection

processes that constructed it piece by piece until it came to characterize the species. Thus, there is no single EEA that can be localized at a particular point in time and space. The EEA will differ for each adaptation and is best described as a statistical aggregate of selection pressures over a particular period of time that are responsible for the emergence of an adaptation (Tooby & Cosmides, 1992).

The hallmarks of adaptation are features that define *special design*—complexity, economy, efficiency, reliability, precision, and functionality (Williams, 1966). These qualities are conceptual criteria subject to empirical testing and potential falsification for any particular hypothesis about an adaptation. Because, in principle, many alternative hypotheses can account for any particular constellation of findings, a specific hypothesis that a feature is an adaptation is, in effect, a probability statement that it is highly unlikely that the complex, reliable, and functional aspects of special design characterizing the feature could have arisen as an incidental by-product of another characteristic or by chance alone (Tooby & Cosmides, 1992). As more and more functional features suggesting special design are documented for a hypothesized adaptation, each pointing to a successful solution to a specific adaptive problem, the alternative hypotheses of chance and incidental by-product become increasingly improbable.

Although adaptations are the primary products of the evolutionary process, they are not the only products. The evolutionary process also produces by-products of adaptations as well as a residue of noise. By-products are characteristics that do not solve adaptive problems and do not have to have functional design. They are carried along with characteristics that do have functional design because they happen to be coupled with those adaptations. The whiteness of bones, for example, is an incidental by-product of the fact that they contain large amounts of calcium, which was presumably selected because of properties such as strength rather than because of whiteness (see Symons, 1992).

An example from the domain of humanly designed artifacts illustrates the concept of a by-product. Consider a particular lightbulb designed for a reading lamp; this lightbulb is designed to produce light. Light production is its function. The design features of a lightbulb—the conducting filament, the vacuum surrounding the filament, and the glass encasement—all contribute to the production of light and are part of its functional design. Lightbulbs also produce heat, however. Heat is a by-product of light production. It is carried along not because the bulb was designed to produce heat but rather because heat tends to be a common incidental consequence of light production.

A naturally occurring example of a by-product of adaptation is the human belly button. There is no evidence that the belly button, per se, helped human ancestors to survive or reproduce. A belly button is not good for catching food, detecting predators, avoiding snakes, locating good habitats, or choosing mates. It does not seem to be involved directly or indirectly in the solution to an adaptive problem. Rather, the belly button is a by-product of something that is an adaptation, namely, the umbilical cord that formerly provided the food supply to the growing fetus. As this example illustrates, establishing the hy-

pothesis that something is a by-product of an adaptation generally requires the identification of the adaptation of which it is a by-product and the reason it is coupled with that adaptation (Tooby & Cosmides, 1992). In other words, the hypothesis that something is a by-product, just like the hypothesis that something is an adaptation, must be subjected to rigorous standards of scientific confirmation and potential falsification. As we discuss below, incidental by-products may come to have their own functions or may continue to have no evolved function at all, and they may be ignored or valued and exploited by people in various cultures.

The third and final product of the evolutionary process is noise, or random effects. Noise can be produced by mutations that neither contribute to nor detract from the functional design of the organism. The glass encasement of a lightbulb, for example, often contains perturbations from smoothness due to imperfections in the materials and the process of manufacturing that do not affect the functioning of the bulb; a bulb can function equally well with or without such perturbations. In self-reproducing systems, these neutral effects can be carried along and passed down to succeeding generations, as long as they do not impair the functioning of the mechanisms that are adaptations. Noise is distinguished from incidental by-products in that it is not linked to the adaptive aspects of design features but rather is independent of such features.

In summary, the evolutionary process produces three products: naturally selected features (adaptations), by-products of naturally selected features, and a residue of noise. In principle, the component parts of a species can be analyzed, and empirical studies can be conducted to determine which of these parts are adaptations, which are by-products, and which represent noise. Evolutionary scientists differ in their estimates of the relative sizes of these three categories of products. Some argue that many obviously important human qualities, such as language, are merely incidental by-products of large brains (e.g., Gould, 1991). Others argue that qualities such as language show evidence of special design that render it highly improbable that it is anything other than a well-designed adaptation for communication and conspecific manipulation (Pinker, 1994). Despite these differences among competing scientific views about the importance and prevalence of adaptations and by-products, all evolutionary scientists agree that there are many constraints on optimal design.

Constraints on Optimal Design

Adaptationists are sometimes accused of being *panglossian*, a term named after Voltaire's (1759/1939) Pangloss, who proposed that everything was for the best (Gould & Lewontin, 1979). According to this criticism, adaptationists are presumed to believe that selection creates optimal design, and practitioners are presumed to liberally spin adaptationist stories. Humans have noses designed to hold up eyeglasses and laps designed to hold computers, and they grow bald so that they can be more easily spotted when lost! This sort of fanciful storytelling, lacking rigorous standards for hypothesis formulation and evidentiary evaluation, would be poor science indeed. Although some no doubt succumb to this sort of cocktail banter, evolutionists going back to Darwin have

long recognized important forces that prevent selection from creating optimally designed adaptations (see Dawkins, 1982, for an extensive summary of these constraints).

First, evolution by selection is a slow process, so there will often be a lag in time between a new adaptive problem and the evolution of a mechanism designed to solve it. The hedgehog's antipredator strategy of rolling into a ball is inadequate to deal with the novel impediment to survival created by automobiles. The moth's mechanism for flying toward light is inadequate for dealing with the novel challenge to survival of candle flames. The existence in humans of a preparedness mechanism for developing a fear of snakes may be a relic not well designed to deal with urban living, which currently contains hostile forces far more dangerous to human survival (e.g., cars, electrical outlets) but for which humans lack evolved mechanisms of fear preparedness (Mineka, 1992). Because of these evolutionary time lags, humans can be said to live in a modern world, but they are burdened with a Stone Age brain designed to deal with ancient adaptive problems, some of which are long forgotten (Allman, 1994).

A second constraint on adaptation occurs because of local optima. A better design may be available, in principle, atop a "neighboring mountain," but selection cannot reach it if it has to go through a deep fitness valley to get there. Selection requires that each step and each intermediate form in the construction of an adaptation be superior to its predecessor form in the currency of fitness. An evolutionary step toward a better solution would be stopped in its tracks if that step caused too steep a decrement in fitness. Selection is not like an engineer who can start from scratch and build toward a goal. Selection works only with the available materials and has no foresight. Local optima can prevent the evolution of better adaptive solutions that might, in principle, exist in potential design space (Dennett, 1995; Williams, 1992).

Lack of available genetic variation imposes a third constraint on optimal design. In the context of artificial selection, for example, it would be tremendously advantageous for dairy breeders to bias the sex ratio of offspring toward milk-producing females rather than nonlactating males. But all selective-breeding attempts to do this have failed, presumably because cattle lack the requisite genetic variation to bias the sex ratio (Dawkins, 1982). Similarly, it might, in principle, be advantageous for humans to evolve X-ray vision to see what is on the other side of obstacles or telescopic vision to spot danger from miles away. But the lack of available genetic variation, along with other constraints, has apparently precluded such adaptations.

A fourth constraint centers on the costs involved in the construction of adaptations. At puberty, male adolescents experience a sharply elevated production of circulating plasma testosterone. Elevated testosterone is linked to onset of puberty, an increase in body size, the production of masculine facial features, and the commencement of sexual interest and activity. But elevated testosterone also has an unfortunate cost—it compromises the immune system, rendering men more susceptible than women to a variety of diseases (Folstad & Karter, 1992; Wedekind, 1992). Presumably, averaged over all men through many generations, the benefits of elevated testosterone outweighed its costs in

the currency of fitness. It evolved despite these costs. The key point is that all adaptations carry costs—sometimes minimal metabolic costs and at other times large survival costs—and these costs impose constraints on the optimal design of adaptations.

A fifth class of constraints involves the necessity of coordination with other mechanisms. Adaptations do not exist in a vacuum, isolated from other evolved mechanisms. Selection favors mechanisms that coordinate well with, and facilitate the functioning of, other evolved mechanisms. This process of coordination, however, often entails compromises in the evolution of an adaptation that render it less efficient than might be optimal in the absence of these constraints. Women, for example, have been selected both for bipedal locomotion and for the capacity for childbirth. The widened hips and birth canal that facilitate childbirth, however, compromise the ability to locomote with great speed. Without the need to coordinate design for running with design for childbirth, selection may have favored slimmer hips like those found on men, which facilitate running speed. The departure from optimal design for running speed in women, therefore, presumably occurs because of compromises required by the need to coordinate adaptive mechanisms with each other.[3] Thus, constraints imposed by the coordination of evolved mechanisms with each other produce design that is less than might be optimal if the mechanisms were not required to coexist.

Time lags, local optima, lack of available genetic variation, costs, and limits imposed by adaptive coordination with other mechanisms all constitute some of the major constraints on the design of adaptations, but there are others (Dawkins, 1982; Williams, 1992). Adaptations are not optimally designed mechanisms. They are better described as jerry-rigged, meliorative solutions to adaptive problems constructed out of the available materials at hand, constrained in their quality and design by a variety of historical and current forces.

Exaptations and Spandrels

Recently, Stephen J. Gould (1991, l997b; see also Gould & Lewontin, 1979; Gould & Vrba, 1982) proposed that the concept of exaptation is a crucial tool for evolutionary psychology, providing a critical supplement to the concept of adaptation. According to this argument, some evolutionary biologists and psychologists have conflated the historical origins of a mechanism or structure with its current utility. For example, the feathers of birds may have originated as evolved mechanisms for thermal regulation. Over evolutionary time, however, the feathers appear to have been co-opted for a different function— flight. According to this distinction, the term *adaptation* would be properly applied to the original thermal regulalion structure and function, but the term *exaptation* would be more appropriate for describing the current flight-producing structure and function.

Gould (1991) provided two related definitions of exaptations. First, an exaptation is "a feature, now useful to an organism, that did not arise as an adaptation for its present role, but was subsequently co-opted for its current function" (p. 43). Second, exaptations are "features that now enhance fitness,

but were not built by natural selection for their current role" (p. 47). On the basis of these related definitions, a mechanism must have a function and must enhance the fitness of its bearer to qualify as an exaptation.

It should be noted that Gould was inconsistent in his usage of the concept of exaptation, even within a single article (e.g., Gould, 1991). Although the definitions of exaptation quoted verbatim here appear to reflect his most common usage (indeed, the quoted 1991 definition was first introduced by Gould and Vrba in 1982), at other times, he seemed to use the term to cover novel but functionless uses or consequences of existing characteristics. For conceptual clarity, it is critical to distinguish between exaptation, as Gould (1991) defined it in the quoted passages, and by-products that are unrelated to function in the biological sense. In the next section, we examine Gould's various usages of the term *exaptation*. However, in this article, we use *exaptation*, consistent with the above quoted definitions, to refer only to mechanisms that have new biological functions that are not the ones that caused the original selection of the mechanisms. Biologically functionless uses are referred to as "effects," "consequences," or "by-products." These two easily confused strands of Gould's discussion of exaptation are thus disentangled here and treated separately.

According to Gould (1991), exaptations come in two types. In the first type, features that evolved by selection for one function are co-opted for another function. We use the term *co-opted adaptation* to describe this first category. The feathers of birds first having evolved for thermal regulation but then later co-opted for flight is an example of a co-opted adaptation. In the second type, "presently useful characteristics did not arise as adaptations ... but owe their origin to side consequences of other features" (Gould, 1991, p. 53). Gould called such side effects of the organism's architecture "spandrels." The term *spandrels* is an architectural term that refers to the spaces left over between structural features of a building. The spaces between the pillars of a bridge, for example, can subsequently be used by homeless persons for sleeping, even though such spaces were not designed for providing such shelter.

In sum, Gould (1991) proposed two types of functional exaptations—adaptations that initially arose through natural selection and were subsequently co-opted for another function (co-opted adaptations) and features that did not arise as adaptations through natural selection but rather as side effects of adaptive processes and that have been co-opted for a biological function (co-opted spandrels). In both cases, according to Gould's primary definition, a mechanism must possess a biological function that contributes to fitness to qualify as an exaptation.

As an example of an exaptation, Gould (1991) used the large size of the human brain and its function of enabling humans to produce speech. The large brain size, according to his argument, originally arose as an adaptation for some (unspecified) functions in humans' ancestral past (Gould, 1991). But the complexity of the human brain produces many by-products that are not properly considered to be functions of the brain: "The human brain, as nature's most complex and flexible organ, throws up spandrels by the thousands for each conceivable adaptation in its initial evolutionary restructuring" (Gould, 1991, p. 58). Among the spandrels he cited as being by-products of large brains are religion, reading, writing, fine arts, the norms of commerce, and the practices of

war. These seem to be intended as functionless uses or by-products rather than true fitness-enhancing, co-opted spandrels. Gould (1991) concluded that among features of interest to psychologists, such by-products are "a mountain to the adaptive molehill" (p. 59).

From these arguments, Gould (1991) concluded that the concepts of exaptations and spandrels provide a "one-line refutation of ... an ultra-Darwinian theory based on adaptation" (p. 58). The two standard pillars of evolutionary biology—natural selection and adaptation—cannot, in principle, account for human behavior "without fatal revisions in its basic intent" (p. 58). Note that Gould was not challenging the importance of evolutionary biology for understanding human behavior. Indeed, as we show later in this article, understanding the nature of the adaptation responsible for producing spandrels (in this case, the nature of the large human brain) is critical to the analysis. Rather, he argued that there has been an overreliance on explanation in terms of adaptation, and to this important explanatory concept must be added the concept of exaptation, which is "a crucial tool for evolutionary psychology" (Gould, 1991, p. 43).

Terminological and Conceptual Confusions in the Invocation of Exaptation and Adaptation

To apply evolutionary concepts to psychology and to properly evaluate and contrast the concepts of exaptation and adaptation as potentially critical tools for evolutionary psychology, several distinctions need to be made, and some common terminological confusions should be clarified.

Confusion 1: Adaptation versus Intuitions about Psychological Adjustment

Psychologists often use the term *adaptive* or *maladaptive* in a colloquial nonevolutionary sense. Often, these usages refer to notions such as personal happiness, social appropriateness, the ability to adjust to changing conditions, or other intuitive notions of well-being. It is important to distinguish these colloquial uses from the technical evolutionary uses, although evolved mechanisms may eventually turn out to be important in explaining personal happiness, well-being, or the ability to adjust to changing conditions (see, e.g., Nesse, 1990).

Confusion 2: Current Utility versus Explanation in Terms of Past Functionality

Taken literally, Gould's (1991) cited definition of exaptation requires that a feature be co-opted for its current function and that it now enhances fitness. It may seem from these phrases that exaptations concern only functions operating at the present moment, whether or not they operated in the past. However, evolutionary psychologists and biologists are generally interested in explaining existing features of organisms. Obviously, a characteristic cannot be explained by current fitness-enhancing properties that came about after the characteristic already existed. When evolutionists attempt to explain the existence of a feature, they must do so by reference to its evolutionary history. All evolutionary explanations of the existence of species-wide mechanisms are to this extent explanations in terms of the past fitness effects of that kind of mechanism that

led to the current existence of the mechanism in the species. The fact that a mechanism currently enhances fitness, by itself, cannot explain why the mechanism exists or how it is structured (Tooby & Cosmides, 1990b).

There are good reasons to think that it is not scientifically illuminating to demonstrate a feature's current correlation with fitness (Symons, 1992; Tooby & Cosmides, 1990b), unless such correlations reveal longer term, past selective pressures. It is not clear that such correlations shed any light on the mechanism's design or status as an adaptation. Such correlations may reveal the current direction of selection, although even this assumes that such correlations will continue to be obtained in future generations—a questionable assumption given the rapidly changing biotic and abiotic environments. Evolutionary explanation focuses on explaining why a feature exists, not what incidental interactions the feature may be having with the current environment.

Confusion 3: Current Functions versus Past Functions That Are No Longer Active

Another confusion lurking in Gould's (1991) language is that it seems to imply that the past functions that explain the existence of a mechanism must still be operating now and literally be a current function to be an adaptation or exaptation. The concepts of adaptation and exaptation are intended as explanatory concepts, and they may be explanatorily useful even when the cited functions are no longer operative. Selected features often cease having the fitness-enhancing effects that got them selected in the first place; for example, it is possible that a selected taste for fatty foods to ensure adequate caloric intake is no longer fitness-enhancing in industrial societies where excessive fat is harmfully common and available for consumption. When evolutionists attempt to explain why humans have a taste for fatty foods, however, they generally say that this taste likely is (or was) an adaptation to ensure adequate caloric intake. Current fitness enhancement is not at issue; at issue is the past function explaining the existence of the mechanisms behind the taste for fatty foods.

A similar point holds for an exaptation. For example, if birds that fly subsequently were to become nonflying, so their feathers would no longer have the exapted function of supporting flight, the existence of feathers at that future time would still need to be explained in terms of (a) an original adaptation for heat insulation and (b) a later exaptation for flying, followed by (c) a functionless period too short for feathers to be selected out. So, the use of exaptation as an evolutionary explanatory concept does not require that there be a current function, any more than the use of adaptation requires such a current function. However, the use of exaptation requires, as Gould (1991) was trying to convey, that there be an original function and a distinct later function (he appeared to use "current" to conveniently distinguish the later function from the original function). What is required for exaptational explanation is not that there be an active current function but that there was an active function at the time that the feature is claimed to have served as an exaptation.

Confusion 4: Function versus Functionless By-Product

The most central confusion in applying Gould's (1991) ideas pertains to distinguishing between exaptations, as Gould defined them, and the novel use of existing features that are currently unrelated to function and fitness. Al-

though Gould (1991) defined an exaptation as a feature "coopted for its current function" (p. 43) and features that "now enhance fitness, but were not built by natural selection for their current role" (p. 46), he sometimes argued that "function" does not describe the utility of exaptations; instead, he suggested that the utility of an exaptation is better described as "effect" (p. 48). Even more confusing, he referred to "culturally useful features" (p. 58) of the brain as exaptations. Gould's stated definitions seem to require that these effects and culturally useful features must contribute to fitness and have specifiable biological functions to qualify as exaptations, but it seems implausible that Gould intended to claim that such cultural practices as reading and writing are explainable by biological functions. Accordingly, exaptations must be distinguished from novel uses of existing mechanisms, where the novel uses are not explained by a biological function.

Consider the human hand as an adaptation. Clearly, the human hand is now used for many activities that were not part of its original set of functions—playing handball or disc golf, manipulating a joystick on a Super Nintendo game, or writing a computer program by pecking on a keyboard. But it seems unlikely that Gould (1991) meant to claim that these activities serve any functions in the formal sense, as solutions to adaptive problems that contribute to reproduction, although they certainly serve *functions* in the colloquial meaning of the term—helping to achieve some goal (e.g., staying in shape, engaging in a stimulating and distracting activity). The same problem arises for many of the activities enumerated by Gould as hypothesized exaptations of the large human brain. Indeed, many of the features Gould claimed to be exaptations or spandrels in human behavior do not seem to fall under his own definitions of exaptation or spandrel and seem instead to be functionless by-products. The key point is that novel uses of existing mechanisms that are not explained by biological function or fitness (i.e., functionless by-products) must be distinguished from true functional exaptations, such as the feathers of birds co-opted for flight.

Confusion 5: What Causal Process or Mechanism Is Doing the Co-opting?

Intimately related to the confusion between exaptations and functionless by-products is a confusion pertaining to the causal process responsible for co-opting an existing structure (see Pinker, 1997a). In the example of birds' feathers, which were originally evolved for thermal regulation but subsequently co-opted for flight, it is clearly natural selection that is responsible for transforming an existing structure into a new, modified structure with a different function. In other cases, however, Gould (1991) appeared to imply that human psychological capacities, such as cognitive capacities, human instrumental actions, or motivational mechanisms, are responsible for the co-opting.

The distinction that evolutionary psychologists make between underlying mechanisms and manifest behavior is helpful in clarifying this confusion. Both adaptations and exaptations, as underlying mechanisms, may be subsequently used for novel behaviors that may have no functional relevance whatsoever. When people use their hands to grip a tennis racquet, for example, this evolutionarily recent manifest behavior is clearly not the function for which the hands evolved. A full understanding of this novel behavior, however, requires

an understanding of the underlying mechanism that is used (the hand) and is aided by insight into the functions for which it was designed (e.g., the power grip). The activity (e.g., tennis) may be partially understood by invoking evolved motivational mechanisms (e.g., social networking, hierarchy negotiation, enhancement of appearance) that are responsible for humans co-opting or exploiting existing mechanisms to pursue this novel activity.

In this example, human motivational mechanisms conjoined with current cognitive and physical capacities, not natural selection, are responsible for co-opting the existing mechanism of the hand. The same logic applies to many of Gould's (1991) other examples of exaptations, such as reading and writing—these are evolutionarily novel activities that are presumably too recent to have been co-opted by natural selection and so apparently must have been invented and co-opted by existing human psychological mechanisms. Such human co-optation must be distinguished from biological exaptations that natural selection has transformed from one function to another.

In summary, evolutionary functional analysis is useful regardless of whether natural selection or some other causal process, such as an existing human motivation, is responsible for the co-opting. Even in cases where a feature has no biological function and is proposed to be a functionless by-product, an understanding of novel behaviors must involve (a) an understanding of the evolved mechanisms that make humans capable of performing the behavior and (b) an understanding of the evolved cognitive and motivational mechanisms that led humans to exploit such capabilities. It is not sufficient from a scientific point of view to merely present a long speculative list of purported exaptations, however interesting or intuitively compelling they might be.

The hypothesis that something is an exaptation or even a functionless effect should be subjected to reasonable standards of hypothesis formulation and empirical verification, just as hypotheses about adaptation must meet these standards. The hypothesis that religion, to use one of Gould's (1991) examples, is an exaptation would seem to require a specification of (a) the original adaptations or by-products that were co-opted to produce religion; (b) the causal mechanism responsible for the co-opting (e.g., natural selection or an existing motivational mechanism); and (c) the exapted biological function of religion, if any; that is, the manner in which it contributes to the solution to an adaptive problem of survival or reproduction. These predictions can then be subjected to evidentiary standards of empirical testing and potential falsification.

Hypotheses about functionless by-products must meet rigorous scientific standards that include a functional analysis of the original adaptations responsible for producing the functionless by-products and the existing human cognitive and motivational mechanisms responsible for the co-opting. Without this specification, the mere assertion that this or that characteristic is an exaptation encounters the same problem that Gould (1991) leveled against adaptationists—the telling of "just-so stories."

Confusion 6: Are Exaptations Merely Adaptations?

A final conceptual issue pertains to whether the concept of exaptation is usefully distinct from the concept of adaptation. Dennett (1995) argued that it is not:

> According to orthodox Darwinism, every adaptation is one sort of exaptation or the other—this is trivial, since no function is eternal; if you go back far enough, you will find that every adaptation has developed out of predecessor structures each of which either had some other use or no use at all. (p. 281)

If all adaptations are exaptations, and all exaptations are adaptations, then having two terms to describe one thing would certainly be superfluous.

Although Dennett's (1995) argument has some merit in pointing to the limits of the distinction between adaptation and exaptation, we think he is wrong in suggesting that there is no difference, and we believe that there is utility in differentiating between the two concepts. Granted, the distinction may end up being more a matter of degree than an absolute distinction because exaptations themselves often involve further adaptations; nonetheless, understanding the degree to which a new function is superimposed on a predecessor structure that already existed as an adaptation or as a by-product may indeed shed light on its nature. The notion that a bird's feathers originally were designed for thermal regulation rather than for flying, for example, may help to explain some of its current features that do not seem to contribute to flight (e.g., insulating, heat-retention features).

In sum, Gould's (1991) concept of exaptation can be meaningfully distinguished from adaptation. Both concepts invoke function; therefore, both must meet the conceptual and evidentiary standards for invoking function. The concepts differ, however, in that adaptations are characteristics that spread through the population because they were selected for some functional effect, whereas exaptations are structures that already exist in the population and continue to exist, albeit sometimes in modified form, for functional reasons different from the ones for which they were originally selected.

The Role of Natural Selection in Adaptations and Exaptations

Some readers of Gould (1997a) come away believing that the role of natural selection is somehow diminished to the degree that exaptations are important. This is a mistake, as Gould himself took pains to point out: "I accept natural selection as the only known cause of 'eminently workable design' and ... 'adaptive design must be the product of natural selection'" (p. 57). Natural selection plays a key role in both adaptations and exaptations.

When exaptations are co-opted adaptations, where the mechanism being co-opted for a new function was an adaptation, selection is required to explain the original adaptation being co-opted. Fishes' fins designed for swimming may have been co-opted to produce mammalian legs for walking. Birds' feathers, perhaps originally designed for thermal regulation, may have been co-opted for flying. In all these cases, however, natural selection is required to explain the origins and nature of the adaptations that provided the existing structures capable of being co-opted.

When exaptations are co-opted spandrels, where the mechanism being co-opted for a new function was not an adaptation but rather an incidental by-product of an adaptation, then selection is required to explain the adaptation

that produced the incidental by-product. Recall that the hypothesis that a mechanism with a function is a spandrel implies that the mechanism was a by-product, and supporting a by-product hypothesis generally requires specifying the adaptation responsible for producing the by-product (Tooby & Cosmides, 1992). Natural selection is required to explain the origin and design of the adaptation—it is the only known causal process capable of producing adaptation. Without specifying the origin of the adaptation that produced the by-product that was co-opted to become a spandrel, the hypothesis that something is a spandrel generally cannot be tested.

Selection is necessary not only to explain the adaptations and by-products that are available for co-optation but also to explain the process of exaptation itself. Selection is required to explain the structural changes in an existing mechanism that enable it to perform the new exapted function: "Exaptations almost always involve structural changes that enable the preexisting mechanism, designed for another function, to perform the new function; these changes require explanation by natural selection" (Wakefield, 1999). When feathers for thermal regulation become wings capable of flight, it is highly unlikely that the new function can occur without any modification of the original mechanism. Selection would have to act on the existing feathers, favoring those individuals that possess more aerodynamic features over those possessing less aerodynamic features. Furthermore, these changes would have to be coordinated with other changes, such as a musculature capable of generating sufficient flapping, alterations in the visual system to accommodate the new demands of aerial mobility, and perhaps modifications of the feet to facilitate landing without damage (e.g., a redesigned shape of the feet). All these changes require the invocation of natural selection to explain the transformation of the original adaptation to an exaptation (e.g., an adaptation with a new function). Similar explanations would generally be necessary for explaining how functionless by-products are transformed into co-opted spandrels that perform specific functions.

Selection is also required to explain the maintenance of an exaptation over evolutionary time, even if no changes in structure occur: "Even in rare cases where exaptations involve no structural changes whatsoever, selective pressures must be invoked to fully explain why the mechanism is maintained in the population" (Wakefield, 1999). The forces of selection, of course, are never static. The fact that more than 99% of all species that have ever existed are now extinct is harsh testimony to the changes in selection over time (Thiessen, 1996). If the selection pressure responsible for the original adaptation becomes neutral or reversed, then the adaptation will eventually degrade over time because of forces such as the cumulative influx of new mutations and competing metabolic demands of other mechanisms. Selection is not only the force responsible for the origins of complex mechanisms but also the force responsible for their maintenance. Thus, even in the odd event that an existing mechanism is co-opted for a new function with no change whatsoever, selection is required to explain why this mechanism and its new function are maintained in the population over time.

In summary, adding exaptation to the conceptual toolbox of evolutionary psychology does not diminish the importance of natural selection as the pri-

mary process responsible for creating complex organic design—a point apparently endorsed by all sides involved in these conceptual debates. Selection is responsible for producing the original adaptations that are then available for co-optation. It is responsible for producing the adaptations, of which spandrels are incidental by-products. It is responsible for producing structural changes in exaptations in order to fulfill their new functions. And it is responsible for maintaining exaptations in the population over evolutionary time, even in the rare cases where no structural changes occurred. The distinctions between exaptation and adaptation are important, and Gould (1991) deserves credit for highlighting them. However, the distinctions should not be taken to mean that natural selection is not the basic explanatory principle in biology and evolutionary psychology.

Testing Hypotheses about Adaptations, Exaptations, and Spandrels

Evolutionary psychological hypotheses about adaptations are sometimes derided as mere storytelling, but the same accusation can be leveled at hypotheses about exaptations and spandrels, and even at more standard social science notions such as socialization, learning, and culture as causal explanations (Tooby & Cosmides, 1992). In all these approaches, as in the case of evolutionary hypotheses about adaptation, it is easy to concoct hypotheses about how a feature might be explained. The key issue is not whether a hypothesis is a story or not—at some level, all scientific hypotheses can be viewed as stories. Rather, the key questions are (a) Is the evolutionary psychological hypothesis formulated in a precise and internally consistent manner? (b) Does the hypothesis coordinate with known causal processes in evolutionary biology, much as hypotheses in cosmology must coordinate with known laws of physics? (Tooby and Cosmides [1992] called this "conceptual integration") (c) Can new specific empirical predictions about behavior or psychology be derived from the hypothesis for which data are currently lacking? (d) Can the hypothesis more parsimoniously account for known empirical findings, and overall, is it more evidentially compelling than competing hypotheses? and (e) Is the proposed psychological mechanism computationally capable of solving the hypothesized problem (Cosmides & Tooby, 1994; Marr, 1982)? These are scientific criteria that can be applied whether the hypothesis is or is not explicitly evolutionary and whether the hypothesis invokes an adaptation, exaptation, spandrel, or functionless by-product.

There is nothing about the fact that a hypothesis is explicitly evolutionary that makes it virtuous or more likely to be correct. Many evolutionarily inspired hypotheses turn out to be wrong, however reasonable they may seem. The hypothesis that the female orgasm functions to facilitate sperm transport, for example, is eminently reasonable on evolutionary grounds and leads to specific testable predictions. At present, however, the evidence for this hypothesis is weak (Baker & Bellis, 1995). In contrast, the hypothesis that male sexual jealousy has evolved to serve the function of combating paternity uncertainty has accrued a reasonable volume of empirical support across diverse methods, samples, and cultures (Baker & Bellis, 1995; Buss, 1988; Buss et al., 1992; Buss & Shackelford, 1997; Buunk, Angleitner, Oubaid, & Buss, 1996; Daly

& Wilson, 1988; Daly, Wilson, & Weghorst, 1982; Shackelford & Buss, 1996; Symons, 1979; Wiederman & Allgeier, 1993; Wilson & Daly, 1992).

When a particular hypothesis about an evolved mechanism fails to be supported empirically, then a number of options are available to researchers. First, the hypothesis may be right but may have been tested incorrectly. Second, the hypothesis may be wrong, but an alternative functional hypothesis could be formulated and tested. Third, the phenomenon under examination might not represent an adaptation or exaptation at all but might instead be an incidental by-product of some other evolved mechanism, and this hypothesis could be tested.

Researchers then can empirically test these alternatives. Suppose, for example, that the sperm transport hypothesis of the female orgasm turned out to be wrong, with the results showing that women who had orgasms were no more likely to conceive than were women who did not have orgasms. The researchers could first scrutinize the methodology to see whether some flaw in the research design may have gone undetected (e.g., had the researchers controlled for the ages of the women in the two groups, because inadvertent age differences may have concealed the effect?). Second, the researchers could formulate an alternative hypothesis—perhaps the female orgasm functions as a mate selection device, providing a cue to the woman about the quality of the man or his investment in her (see Rancour-Laferriere, 1985, for a discussion of this and other hypotheses about the female orgasm)—and this alternative could be tested. Third, the researchers could hypothesize that the female orgasm is not an adaptation at all but rather an incidental by-product of some other mechanism, such as a common design shared with men, who do possess the capacity for orgasm for functional reasons (see Symons, 1979, for the original proposal of this functionless by-product hypothesis, and Gould's, 1987, subsequent endorsement of this hypothesis). In this case, researchers could try to disconfirm all existing functional explanations and could try to identify how the known mechanisms for development of naturally selected male orgasmic capacities led to the female orgasmic capacities as a side effect. Different researchers undoubtedly will have different proclivities about which of these options they pursue. The key point is that all evolutionary hypotheses—whether about adaptations, exaptations, spandrels, or functionless by-products—should be formulated in a precise enough manner to produce empirical predictions that can then be subjected to testing and potential falsification.

It should be noted that evolutionary hypotheses range on a gradient from well-formulated, precise deductions from known evolutionary principles on the one hand to evolutionarily inspired hunches on the other (see, e.g., Symons, 1992). Evolutionary psychology often provides a heuristic, guiding scientific inquiry to important domains that have a priori importance, such as events surrounding reproduction (e.g., sexuality, mate selection). Just as with a precise evolutionary hypothesis, an evolutionary hunch may turn out to be right or wrong. It would seem reasonable to hypothesize, for example, that men would have evolved mechanisms designed to detect when women ovulate, because such a mechanism would help to solve the adaptive problems of identifying fecund women and channeling mating effort more efficiently. But there is little solid empirical evidence that such a mechanism exists (see Symons, 1995). Such

hunches, however, can often be useful in guiding investigations. Thus, evolutionary psychology, at its best, has both heuristic and predictive value for psychological science.

Discussion

In principle, we agree with Gould's (1991, 1997b) suggestion to be pluralistic about the conceptual tools of evolutionary psychology, although it is clear that many evolutionary psychologists already embody the pluralism advocated (e.g., Tooby & Cosmides, 1990a, 1992). Researchers may differ about which of these tools they believe are most scientifically valuable for particular purposes. One reasonable standard for judging the value of such conceptual tools is the heuristic and predictive empirical harvest they yield. Table 28.1 shows 30 recent examples of the empirical findings about humans whose discovery was guided by hypotheses anchored in adaptation and natural selection.

From this empirical evidence, hypotheses about adaptations appear to have considerable value. In some cases, adaptation-minded researchers have generated and tested specific empirical predictions not generated from nonadaptationist theories, such as sex-linked causes of divorce (Betzig, 1989), causes of the intensity of mate retention effort (Buss & Shackelford, 1997), predictable conditions under which spousal homicide occurs (Daly & Wilson, 1988), sex differences in the nature of sexual fantasy (Ellis & Symons, 1990), and shifts in mate preferences across the life span (Kenrick & Keefe, 1992). In other cases, adaptation-mindedness has proved heuristic, guiding researchers to important domains not previously examined or discovered, such as the role of symmetry in mate attraction (Thornhill & Gangestad, 1993), the role of deception in mate attraction (Tooke & Camire, 1991), and the specific conflicts of interest that occur in stepfamilies (Wilson & Daly, 1987). Using the same criterion, we could not find a single example of an empirical discovery made about humans as a result of using the concepts of exaptations or spandrels (but see MacNeilage, 1997, for a testable exaptation hypothesis about the origins of human speech production). Of course, this relative lack of fruitfulness at this time does not imply that over time, the concepts of exaptation and spandrels cannot be useful in generating scientific hypotheses and producing empirical discoveries.

In this article, we have attempted to elucidate the defining criteria of adaptations, exaptations, spandrels, and functionless by-products. Tables 28.2 and 28.3 summarize several important conceptual and evidentiary standards applicable to each of these concepts.

Adaptations and exaptations—in the form of either co-opted adaptations or co-opted spandrels—share several common features. All invoke selection at some point in the causal sequence. All invoke function. All must meet conceptual criteria for the proposed function—the hallmarks of special design, including specialization of function for solving a particular adaptive problem. And all must meet evidentiary standards, such as generating specific testable empirical predictions and parsimoniously accounting for known empirical findings.

These concepts differ, however, in the role of selective origins and fitness in explaining a feature. Although all three invoke selection, adaptations that arose de novo from mutations invoke selection in the original construction of the

Table 28.1
Thirty recent examples of empirical discoveries about humans generated by thinking about adaptation and selection

Example	Source
Evolved landscape preferences	Orions & Heerwagen (1992)
Sexually dimorphic mating strategies	Thiessen (1993); Thiessen, Young, & Burroughs (1993)
Waist-to-hip ratio as a determinant of attractiveness judgments	Singh (1993)
Standards of beauty involving symmetry	Grammer & Thornhill (1994)
Women's desire for mates with resources found in 37 cultures	Buss (1989)
Men's preference for younger mates documented in 37 cultures	Buss (1989)
Cheater detection procedure in social exchange	Cosmides (1989)
Stepchild abuse at 40 times the rate of nonstepchild abuse	Wilson & Daly (1987)
Relationship-specific sensitivity to betrayal	Shackelford & Buss (1996)
Sex-linked shifts in mate preference across the life span	Kenrick & Keefe (1992)
Predictable patterns of spousal and same-sex homicide	Daly & Wilson (1988)
Pregnancy sickness as an adaptation to teratogens	Profet (1992)
Mother–fetus conflict	Haig (1993)
Predictably patterned occurrence of allergies	Profet (1991)
Different human sperm morphs	Baker & Bellis (1995)
Superior female spatial location memory	Silverman & Eals (1992)
Design of male sexual jealousy	Buss et al. (1992); Daly et al. (1982)
Sex differences in sexual fantasy	Ellis & Symons (1990)
Deception in mating tactics	Tooke & Camire (1991)
Profiles of sexual harassers and their victims	Studd & Gattiker (1991)
Sex differences in desire for sexual variety	Clark & Hatfield (1989)
Facial asymmetry as an indicator of poor psychological and physical health	Shackelford & Larsen (1997)
Frequentist reasoning in human cognition	Cosmides & Tooby (1996); Gigerenzer & Hoffrage (1995)
Predictable causes of conjugal dissolution in 89 cultures	Betzig (1989)
Socialization practices across cultures differing by sex and mating system	Low (1989)
Patterns of risk taking in intrasexual competition for mates	Wilson & Daly (1985)
Shifts in grandparental investment according to sex of grandparent and sex of parent	DeKay (1995); Euler & Weitzel (1996)
Perceptual adaptations for entraining, tracking, and predicting animate motion	Heptulla-Chatterjee, Freyd, & Shiffrar (1996)
Universal perceptual adaptations to terrestrial living	Shepard (1984, 1992)
Mate guarding as a function of female reproductive value	Buss & Shackelford (1997); Dickemann (1981)

Table 28.2
Conceptual and evidentiary criteria for evaluating the core concepts of adaptations, exaptations, spandrels, and functionless by-products

Differentiation criteria	Adaptation	Exaptation: Co-opted adaptation	Co-opted spandrel	Functionless by-product
Origin and maintenance	History of selection	Selection operating on previous adaptation	Selection operating on previous by-product	History of selection for mechanism that produced by-product
Role of fitness	Correlated with fitness in past during period of its evolution	Currently correlated with fitness	Currently correlated with fitness	Not directly related to fitness
Critical features	Solved adaptive problem in past	Has new function	Has new function	No previous or current function

Note: Exaptations and *spandrels* are used here according to Gould's (1991) primary meanings, that is, as features co-opted for new current functions; *functionless by-product* is the term used for Gould's other and less common usages of exaptations and spandrels, that is, as incidental, nonfunctional consequences of other characteristics. In the evolutionary literature, these are usually called "by-products." In Gould's usage, "currently enhances fitness" presumably refers to the period of evolutionary time during which selection transformed a previous adaptation or by-product into a new function. Note also that Gould sometimes used the term *exaptation* to cover both co-opted adaptations and co-opted spandrels; we treat these separately.

Table 28.3
Standards common to adaptations, exaptations (co-opted adaptations), and co-opted spandrels

Standards	Criteria
Conceptual	Hallmarks of special design for proposed function: complexity, efficiency, reliability, specificity, capability of solving adaptive problem, and evolvability
Empirical	Capable of generating specific and falsifiable empirical predictions; must account for known data better than alternative hypotheses

mechanism as a species-wide feature. Co-opted adaptations invoke selection in the original construction of the mechanism that is co-opted as well as in any reconstruction necessary for reshaping the mechanism for its new function and in maintaining the mechanism in the population because of its new function. And co-opted spandrels invoke selection in explaining the adaptations of which they are by-products, in explaining the reshaping of the by-product for its new function, and in explaining the maintenance of the by-product in the population because of its new function. Consequently, relative to initial adaptations, exaptations carry the additional evidentiary burden of showing that a current function is distinct from an earlier function or from a functional original structure.

The most important differences, however, center on the temporal aspect of function and fitness. Adaptations exist in the present because their form was shaped in the past by selection for a particular function (Darwin, 1859/1958;

Symons, 1979; Tooby & Cosmides, 1990b; Williams, 1966). Exaptations, in contrast, exist in the present because they were co-opted from previous structures that evolved for reasons different from those of the later exapted function (Gould, 1991). Although all three concepts require documentation of special design for a hypothesized function, co-opted exaptations and spandrels carry the additional evidentiary burdens of documenting both later co-opted functionality and a distinct original adaptational functionality. To our knowledge, none of the items on Gould's (1991) list of proposed spandrels and exaptations—language, religion, principles of commerce, warfare, reading, writing, and fine arts—have met these standards of evidence. Moreover, even if they did meet such standards, this would in no way diminish the need to place such items within an overall evolutionary framework in order to adequately understand and explain them—a point agreed on by all sides of these debates.

Evolutionary psychology is emerging as a promising theoretical perspective within psychology. As with many emerging theoretical perspectives, there is often controversy about the meaning and scientific utility of the new explanatory concepts. Although most psychologists cannot be expected to become steeped in all of the formal complexities of the highly technical discipline of evolutionary theory, we hope that this article will serve as a guide to some of the most theoretically useful core concepts and some of the most interesting controversies within this emerging perspective in psychological science.

Acknowledgments

William Bevan served as action editor for this article.

We thank Rick Arnold, George Bittner, Leda Cosmides, Helena Cronin, Todd DeKay, Randy Diehl, Rob Kurzban, Don Symons, Del Thiessen, and John Tooby for discussions and commentary on the ideas contained in this article.

Notes

1. The empirical application of evolutionary ideas to the study of nonhuman animal behavior, of course, has a long and rich history of success (see Alcock, 1993). Indeed, theory and research emerging from the study of animal behavior have been of great benefit to evolutionary psychology, and comparative psychology continues to inform research about humans (Tooby & Cosmides, 1992). Furthermore, over the past 40 years, ethologists have applied evolutionary functional analysis to manifest human behavior, such as in the study of fixed action patterns (e.g., Lorenz, 1952; Tinbergen, 1951) and universals of facial expression (Ekman, 1973). It was not until the late 1980s, however, that underlying psychological mechanisms, such as those postulated by cognitive psychologists subsequent to the cognitive revolution in psychology, were explored empirically from an evolutionary perspective (e.g., Buss, 1989; Cosmides, 1989).
2. Obviously, the inheritance of selected characteristics and their spread throughout a population are much more complex topics than we can do justice to here; for more extended treatments, see Dawkins (1982), Tooby and Cosmides (1992), and Williams (1966).
3. These and other examples throughout this article are used to illustrate the conceptual points being made and should be regarded at this early stage in the development of evolutionary psychology as hypotheses to be subjected to empirical verification.

References

Alcock, J. (1993). *Animal behavior: An evolutionary approach* (5th ed.). Sunderland, MA: Sinauer.
Allman, W. F. (1994). *The Stone Age present.* New York: Simon & Schuster.
Baker, R. R., & Bellis, M. A. (1995). *Human sperm competition.* London: Chapman & Hall.

Baldwin, J. M. (1894). *Mental development in the child and the race.* New York: Kelly.

Belsky, J., Steinberg, L., & Draper, P. (1991). Childhood experience, interpersonal development, and reproductive strategy: An evolutionary theory of socialization. *Child Development*, *62*, 647–670.

Betzig, L. (1989). Causes of conjugal dissolution: A cross-cultural study. *Current Anthropology, 30*, 654–676.

Buss, D. M. (1988). From vigilance to violence: Tactics of mate retention among American undergraduates. *Ethology and Sociobiology, 9*, 291–317.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12*, 1–49.

Buss, D. M. (1994). *The evolution of desire: Strategies of human mating*. New York: Basic Books.

Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry, 6*, 1–30.

Buss, D. M., Larsen, R. J., Westen, D., & Semmelroth, J. (1992). Sex differences in jealousy: Evolution, physiology, and psychology. *Psychological Science, 3*, 251–255.

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review, 100*, 204–232.

Buss, D. M., & Shackelford, T. K. (1997). From vigilance to violence: Tactics of mate retention in married couples. *Journal of Personality and Social Psychology, 72*, 346–361.

Buunk, A. B., Angleitner, A., Oubaid, V., & Buss, D. M. (1996). Sex differences in jealousy in evolutionary and cultural perspective: Tests from The Netherlands, Germany, and the United States. *Psychological Science, 7*, 359–363.

Clark, R. D., & Hatfield, E. (1989). Gender differences in receptivity to sexual offers. *Journal of Psychology and Human Sexuality, 2*, 39–55.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? *Cognition, 31*, 187–276.

Cosmides, L., & Tooby, J. (1994). Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. *Cognition, 50*, 41–77.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*, 1–73.

Daly, M., & Wilson, M. (1988). *Homicide*. Hawthorne, NY: Aldine de Gruyter.

Daly, M., Wilson, M., & Weghorst, S. J. (1982). Male sexual jealousy. *Ethology and Sociobiology, 3*, 11–27.

Darwin, C. (1958). *On the origin of species by means of natural selection*. New York: New American Library. (Original work published 1859)

Darwin, C. (1981). *The descent of man and selection in relation to sex*. Princeton, NJ: Princeton University Press. (Original work published 1871)

Dawkins, R. (1982). *The extended phenotype*. San Francisco: Freeman.

Dawkins, R. (1996). *Climbing Mount Improbable*. New York: Norton.

DeKay, W. T. (1995, June). *Grandparental investment and the uncertainty of kinship*. Paper presented at the Seventh Annual Meeting of the Human Behavior and Evolution Society, Santa Barbara, CA.

DeKay, W. T., & Buss, D. M. (1992). Human nature, individual differences, and the importance of context: Perspectives from evolutionary psychology. *Current Directions in Psychological Science, 1*, 184–189.

Dennett, D. C. (1995). *Darwin's dangerous idea*. New York: Simon & Schuster.

Dickemann, M. (1981). Paternal confidence and dowry competition: A biocultural analysis of purdah. In R. D. Alexander & D. W. Tinkle (Eds.), *Natural selection and social behavior* (pp. 417–438). New York: Chiron.

Ekman, P. (1973). Cross-cultural studies of facial expression. In P. Ekman (Ed.), *Darwin and facial expression* (pp. 169–222). New York: Academic Press.

Ellis, B. J., & Symons, D. (1990). Sex differences in sexual fantasy: An evolutionary psychological approach. *Journal of Sex Research, 27*, 527–556.

Euler, H. A., & Weitzel, B. (1996). Discriminative grandparental solicitude as reproductive strategy. *Human Nature, 7*, 39–59.

Folstad, I., & Karter, A. J. (1992). Parasites, bright males, and the immunocompetence handicap. *American Naturalist, 139*, 603–622.

Gangestad, S. W., & Simpson, J. A. (1990). Toward an evolutionary history of female sociosexual variation. *Journal of Personality, 58*, 69–96.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Reiew, 102*, 684–704.

Gould, S. J. (1987). Freudian slip. *Natural History, 96*(1), 14–21.

Gould, S. J. (1991). Exaptation: A crucial tool for evolutionary psychology. *Journal of Social Issues, 47*, 43–65.

Gould, S. J. (1997a, October 9). Evolutionary psychology: An exchange. *New York Review of Books, XLIV*, 53–58.

Gould, S. J. (1997b). The exaptive excellence of spandrels as a term and prototype. *Proceedings of the National Academy of Sciences, 94*, 10750–10755.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London B, 205*, 581–598.

Gould, S. J., & Vrba, E. S. (1982). Exaptation: A missing term in the science of form. *Paleobiology, 8*, 4–15.

Grammer, K., & Thornhill, R. (1994). Human facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology, 108*, 233–242.

Haig, D. (1993). Maternal–fetal conflict in human pregnancy. *Quarterly Review of Biology, 68*, 495–532.

Hamilton, W. D. (1964). The genetical evolution of social behavior. *Journal of Theoretical Biology, 7*, 1–52.

Heptulla-Chatterjee, S., Freyd, J. J., & Shiffrar, M. (1996). Configural processing in the perception of apparent biological motion. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 916–929.

James, W. (1962). *Principles of psychology*. New York: Dover. (Original work published 1890)

Jennings, H. S. (1930). *The biological basis of human nature*. New York: Norton.

Kenrick, D. T., & Keefe, R. C. (1992). Age preferences in mates reflect sex differences in human reproductive strategies. *Behavioral and Brain Sciences, 15*, 75–133.

Lilienfeld, S. O., & Marino, L. (1995). Mental disorder as a Roschian concept: A critique of Wakefield's "harmful dysfunction" analysis. *Journal of Abnormal Psychology, 104*, 411–420.

Lorenz, K. Z. (1952). *King Solomon's ring*. New York: Cromwell.

Low, B. S. (1989). Cross-cultural patterns in the training of children: An evolutionary perspective. *Journal of Comparative Psychology, 103*, 313–319.

MacNeilage, P. (1997). What ever happened to articulate speech? In M. C. Corballis & S. Lea (Eds.), *Evolution of the hominid mind*. New York: Oxford University Press.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Mineka, S. (1992). Evolutionary memories, emotional processing, and the emotional disorders. In D. Medin (Ed.), *The psychology of learning and motivation* (Vol. 28). New York: Academic Press.

Morgan, C. L. (1896). *Habit and instinct*. London: Arnold.

Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature, 1*, 261–289.

Orions, G. H., & Heerwagen, J. H. (1992). Evolved response to landscapes. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 555–580). New York: Oxford University Press.

Piattelli-Palmarini, M. (1989). Evolution, selection, and cognition: From "learning" to parameter setting in biology and the study of language. *Cognition, 31*, 1–44.

Pinker, S. (1994). *The language instinct*. New York: Morrow.

Pinker, S. (1997a, October 9). Evolutionary psychology: An exchange. *New York Review of Books, XLIV*, 55–56.

Pinker, S. (1997b). *How the mind works*. New York: Norton.

Pinker, S., & Bloom, P. (1992). Natural language and natural selection. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 451–493). New York: Oxford University Press.

Profet, M. (1991). The function of allergy: Immunological defense against toxins. *Quarterly Review of Biology, 66*, 23–62.

Profet, M. (1992). Pregnancy sickness as adaptation: A deterrent to maternal ingestion of teratogens. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 327–365). New York: Oxford University Press.

Rancour-Laferriere, D. (1985). *Signs of the flesh: An essay on the evolution of hominid sexuality*. New York: Mouton de Gruyter.

Richters, J. E., & Cicchetti, D. (1993). Mark Twain meets *DSM-III–R*: Conduct disorder, development, and the concept of harmful dysfunction. *Development & Psychopathology, 5*, 5–29.

Romanes, G. (1889). *Mental evolution in man: Origin of human faculty*. New York: Appleton.

Sedikedes, C., & Skowronski, J. J. (1997). The symbolic self in evolutionary context. *Personality and Social Psychology Review, 1*, 80–102.

Shackelford, T. K., & Buss, D. M. (1996). Betrayal in mateships, friendships, and coalitions. *Personality and Social Psychology Bulletin, 22*, 1151–1164.

Shackelford, T. K., & Larsen, R. J. (1997). Facial asymmetry as an indicator of psychological, emotional, and physiological distress. *Journal of Personality and Social Psychology, 72*, 456–466.

Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review, 91*, 417–447.

Shepard, R. N. (1992). The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 495–532). New York: Oxford University Press.

Silverman, I., & Eals, M. (1992). Sex differences in spatial abilities: Evolutionary theory and data. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 533–549). New York: Oxford University Press.

Singh, D. (1993). Adaptive significance of female physical attractiveness: Role of waist-to-hip ratio. *Journal of Personality and Social Psychology, 65*, 293–307.

Studd, M. V., & Gattiker, U. E. (1991). The evolutionary psychology of sexual harassment in organizations. *Ethology and Sociobiology, 12*, 249–290.

Symons, D. (1979). *The evolution of human sexuality*. New York: Oxford University Press.

Symons, D. (1987). If we're all Darwinians, what's the fuss about? In C. Crawford, D. Krebs, & M. Smith (Eds.), *Sociobiology and psychology* (pp. 121–146). Hillsdale, NJ: Erlbaum.

Symons, D. (1992). On the use and misuse of Darwinism in the study of human behavior. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 137–159). New York: Oxford University Press.

Symons, D. (1995). Beauty is in the adaptations of the beholder: The evolutionary psychology of human female sexual attractiveness. In P. R. Abramson & S. D. Pinkerton (Eds.), *Sexual nature, sexual culture* (pp. 80–118). Chicago: University of Chicago Press.

Than-Than, Hutton, R. A., Myint-Lwin, Khin-EiHan, Soe-Soe, Tin-Nu-Swe, Phillips, R. E., & Warrell, D. A. (1988). Haemostatic disturbances in patients bitten by Russell's viper (*Vipera russelli siamensis*) in Burma. *British Journal of Haematology, 69*, 513–520.

Thiessen, D. (1993). Environmental tracking by females: Sexual lability. *Human Nature, 5*, 167–202.

Thiessen, D. (1996). *Bittersweet destiny*. New Brunswick, NJ: Transaction.

Thiessen, D., Young, R. K., & Burroughs, R. (1993). Lonely hearts advertisements reflect sexually dimorphic mating strategies. *Ethology and Sociobiology, 14*, 209–229.

Thornhill, R., & Gangestad, S. W. (1993). Human facial beauty: Averageness, symmetry, and parasite resistance. *Human Nature, 4*, 237–270.

Tinbergen, N. (1951). *The study of instinct*. London: Oxford University Press.

Tooby, J., & Cosmides, L. (1990a). On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality, 58*, 17–68.

Tooby, J., & Cosmides, L. (1990b). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology, 11*, 375–424.

Tooby, J., & Cosmides, L. (1992). Psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 19–136). New York: Oxford University Press.

Tooke, J., & Camire, L. (1991). Patterns of deception in intersexual and intrasexual mating strategies. *Ethology and Sociobiology, 10*, 241–253.

Voltaire, F. M. A. (1939). *Candide*. London: Nonesuch Press. (Original work published 1759)

Wakefield, J. C. (1992). The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist, 47*, 373–388.

Wakefield, J. C. (1999). Evolutionary versus prototype analyses of the concept of disorder. *Journal of Abnormal Psychology, 108*, 374–399.

Wedekind, C. (1992). Detailed information about parasites revealed by sexual ornamentation. *Proceedings of the Royal Society of London B, 247*, 169–174.

Wiederman, M. W., & Allgeier, E. R. (1993). Gender differences in sexual jealousy: Adaptationist or social learning explanation? *Ethology and Sociobiology, 14*, 115–140.

Williams, G. C. (1966). *Adaptation and natural selection*. Princeton, NJ: Princeton University Press.

Williams, G. C. (1992). *Natural selection*. New York: Oxford University Press.

Wilson, M., & Daly, M. (1985). Competitiveness, risk-taking, and violence: The young male syndrome. *Ethology and Sociobiology, 6*, 59–73.

Wilson, M., & Daly, M. (1987). Risk of maltreatment of children living with stepparents. In R. J. Gelles & J. B. Lancaster (Eds.), *Child abuse and neglect* (pp. 215–232). Hawthorne, NY: Aldine de Gruyter.

Wilson, M., & Daly, M. (1992). The man who mistook his wife for a chattel. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 289–322). New York: Oxford University Press.

Chapter 29

Toward Mapping the Evolved Functional Organization of Mind and Brain

John Tooby and Leda Cosmides

Nothing in biology makes sense except in the light of evolution.
—T. Dobzhansky

It is the theory which decides what we can observe.
—A. Einstein

Seeing with New Eyes: Toward an Evolutionarily Informed Cognitive Neuroscience

The task of cognitive neuroscience is to map the information-processing structure of the human mind and to discover how this computational organization is implemented in the physical organization of the brain. The central impediment to progress is obvious: The human brain is, by many orders of magnitude, the most complex system that humans have yet investigated. Purely as a physical system, the vast intricacy of chemical and electrical interactions among hundreds of billions of neurons and glial cells defeats any straightforward attempt to build a comprehensive model, as one might attempt to do with particle collisions, geological processes, protein folding, or host-parasite interactions. Combinatorial explosion makes the task of elucidating the brain's computational structure even more overwhelming: There is an indefinitely large number of specifiable inputs, measurable outputs, and possible relationships between them. Even worse, no one yet knows with certainty how computations are physically realized. They depend on individuated events within the detailed structure of neural microcircuitry largely beyond the capacity of current technologies to observe or resolve. Finally, the underlying logic of the system has been obscured by the torrent of recently generated data.

Historically, however, well-established theories from one discipline have functioned as organs of perception for others (e.g., statistical mechanics for thermodynamics). They allow new relationships to be observed and make visible elegant systems of organization that had previously eluded detection. It seems worth exploring whether evolutionary biology could provide a rigorous metatheoretical framework for the brain sciences, as they have recently begun to do for psychology (Shepard, 1984, 1987a, 1987b; Gallistel, 1990; Cosmides and Tooby, 1987; Pinker, 1994, 1997; Marr, 1982; Tooby and Cosmides, 1992).

From chapter 80 in *The New Cognitive Neurosciences*, 2d ed., ed. Michael S. Gazzaniga (Cambridge, MA: MIT Press, 2000), 1167–1178. Reprinted with permission.

Cognitive neuroscience began with the recognition that the brain is an organ designed to process information and that studying it as such would offer important new insights. Cognitive neuroscientists also recognize that the brain is an evolved system, but few realize that anything follows from this second fact. Yet these two views of the brain are intimately related and, when considered jointly, can be very illuminating.

Why Brains Exist

The brain is an organ of computation that was built by the evolutionary process. To say that the brain is an organ of computation means that (1) its physical structure embodies a set of programs that process information, and (2) that physical structure is there *because* it embodies these programs. To say that the brain was built by the evolutionary process means that its functional components—its programs—are there *because* they solved a particular problem-type in the past. In systems designed by natural selection, function determines structure.

Among living things, there are whole kingdoms filled with organisms that lack brains (plants, Monera, fungi). The sole reason that evolution introduced brains into the designs of some organisms—the reason brains exist at all—is because brains performed computations that regulated these organisms' internal processes and external activities in ways that promoted their fitness. For a randomly generated modification in design to be selected—that is, for a mutation to be incorporated by means of a nonrandom process into a species-typical brain design—it had to improve the ability of organisms to solve adaptive problems. That is, the modification had to have a certain kind of effect: It had to improve the organisms' performance of some activity that systematically enhanced the propagation of that modification, summed across the species' range and across many generations. This means that the design of the circuits, components, systems, or modules that make up our neural architecture must reflect, to an unknown but high degree, (1) the computational task demands inherent in the performance of those ancestral activities and (2) the evolutionarily long-enduring structure of those task environments (Marr, 1982; Shepard, 1987a; Tooby and Cosmides, 1992).

Activities that promoted fitness in hominid ancestral environments differ in many ways from activities that capture our attention in the modern world, and they were certainly performed under radically different circumstances. (Consider: hunting vs. grocery shopping; walking everywhere vs. driving and flying; cooperating within a social world of ~200 relatives and friends vs. 50,000 strangers in a medium-sized city). The design features of the brain were built to specifications inherent in ancestral adaptive problems and selection pressures, often resulting in talents or deficits that seem out of place or irrational in our world. A baby cries—alerting her parents—when she is left to sleep alone in the dark, not because hyenas roam her suburban household, but because her brain is designed to keep her from being eaten under the circumstances in which our species evolved.

There is no single algorithm or computational procedure that can solve every adaptive problem (Cosmides and Tooby, 1987; Tooby and Cosmides, 1990a,

1992). The human mind (it will turn out) is composed of many different programs for the same reason that a carpenter's toolbox contains many different tools: Different problems require different solutions. To reverse-engineer the brain, one needs to discover functional units that are native to its organization. To do this, it is useful to know, as specifically as possible, what the brain is for—which specific families of computations it was built to accomplish and what counted as a biologically successful outcome for each problem-type. The answers to this question must be phrased in computational terms because that is the only language that can capture or express the functions that neural properties were naturally selected to embody. They must also refer to the ancestral activities, problems, selection pressures, and environments of the species in question because jointly these define the computational problems each component was configured to solve (Cosmides and Tooby, 1987; Tooby and Cosmides, 1990a, 1992).

For these reasons, evolutionary biology, biological anthropology, and cognitive psychology (when integrated, called *evolutionary psychology*) have the potential to supply to cognitive neuroscientists what might prove to be a key missing element in their research program: a partial list of the native information-processing functions that the human brain was built to execute, as well as clues and principles about how to discover or evaluate adaptive problems that might be proposed in the future.

Just as the fields of electrical and mechanical engineering summarize our knowledge of principles that govern the design of human-built machines, the field of evolutionary biology summarizes our knowledge of the engineering principles that govern the design of organisms, which can be thought of as machines built by the evolutionary process (for overviews, see Daly and Wilson, 1984; Dawkins, 1976, 1982, 1986; Krebs and Davies, 1997). Modern evolutionary biology constitutes, in effect, a foundational "organism design theory" whose principles can be used to fit together research findings into coherent models of specific cognitive and neural mechanisms (Tooby and Cosmides, 1992). To apply these theories to a particular species, one integrates analyses of selection pressures with models of the natural history and ancestral environments of the species. For humans, the latter are provided by hunter–gatherer studies, biological anthropology, paleoanthropology, and primatology (Lee and DeVore, 1968).

First Principles: Reproduction, Feedback, and the Antientropic Construction of Organic Design

Within an evolutionary framework, an organism can be described as a self-reproducing machine. From this perspective, the defining property of life is the presence in a system of "devices" (organized components) that cause the system to construct new and similarly reproducing systems. From this defining property—self-reproduction—the entire deductive structure of modern Darwinism logically follows (Dawkins, 1976; Williams, 1985; Tooby and Cosmides, 1990a). Because the replication of the design of the parental machine is not always error free, randomly modified designs (i.e., mutants) are introduced into populations of reproducers. Because such machines are highly organized

so that they cause the otherwise improbable outcome of constructing offspring machines, most random modifications interfere with the complex sequence of actions necessary for self-reproduction. Consequently, such modified designs will tend to remove themselves from the population—a case of negative feedback.

However, a small residual subset of design modifications will, by chance, happen to constitute improvements in the design's machinery for causing its own reproduction. Such improved designs (by definition) cause their own increasing frequency in the population—a case of positive feedback. This increase continues until (usually) such modified designs outreproduce and thereby replace all alternative designs in the population, leading to a new species-standard design. After such an event, the population of reproducing machines is different from the ancestral population: The population- or species-standard design has taken a step "uphill" toward a greater degree of functional organization for reproduction than it had previously. This spontaneous feedback process—natural selection—causes functional organization to emerge *naturally*, that is, without the intervention of an intelligent "designer" or supernatural forces.

Over the long run, down chains of descent, this feedback cycle pushes designs through state-space toward increasingly well-organized—and otherwise improbable—functional arrangements (Dawkins, 1986; Williams, 1966, 1985). These arrangements are functional in a specific sense: the elements are improbably well organized to cause their own reproduction in the environment in which the species evolved. Because the reproductive fates of the inherited traits that coexist in the same organism are linked together, traits will be selected to enhance each other's functionality (however, see Cosmides and Tooby, 1981, and Tooby and Cosmides, 1990a, for the relevant genetic analysis and qualifications). As design features accumulate, they will tend to sequentially fit themselves together into increasingly functionally elaborated machines for reproduction, composed of constituent mechanisms—called *adaptations*—that solve problems that either are necessary for reproduction or increase its likelihood (Darwin, 1859; Dawkins, 1986; Thornhill, 1991; Tooby and Cosmides, 1990a; Williams, 1966, 1985). Significantly, in species like humans, genetic processes ensure that complex adaptations virtually always are species-typical (unlike nonfunctional aspects of the system). This means that *functional* aspects of the architecture will tend to be universal at the genetic level, even though their expression may often be sex or age limited, or environmentally contingent (Tooby and Cosmides, 1990b).[1]

Because design features are embodied in individual organisms, they can, generally speaking, propagate themselves in only two ways: by solving problems that increase the probability that offspring will be produced either by the organism they are situated in or by that organism's kin (Hamilton, 1964; Williams and Williams, 1957; however, see Cosmides and Tooby, 1981, and Haig, 1993, for intragenomic methods). An individual's relatives, by virtue of having descended from a recent common ancestor, have an increased likelihood of having the same design feature as compared to other conspecifics. This means that a design modification in an individual that causes an increase in the reproductive rate of that individual's kin will, by so doing, tend to increase its

own frequency in the population. Accordingly, design features that promote both direct reproduction and kin reproduction, and that make efficient trade-offs between the two, will replace those that do not. To put this in standard biological terminology, design features are selected to the extent that they promote their inclusive fitness (Hamilton, 1964).

In addition to selection, mutations can become incorporated into species-typical designs by means of chance processes. For example, the sheer impact of many random accidents may cumulatively propel a useless mutation upward in frequency until it crowds out all alternative design features from the population. Clearly, the presence of such a trait in the architecture is not explained by the (nonexistent) functional consequences that it had over many generations on the design's reproduction; as a result, chance-injected traits will not tend to be coordinated with the rest of the organism's architecture in a functional way.

Although such chance events play a restricted role in evolution and explain the existence and distribution of many simple and trivial properties, organisms are not primarily chance agglomerations of stray properties. Reproduction is a highly improbable outcome in the absence of functional machinery designed to bring it about, and only designs that retain all the necessary machinery avoid being selected out. To be invisible to selection and, therefore, not organized by it a modification must be so minor that its effects on reproduction are negligible. As a result, chance properties do indeed drift through the standard designs of species in a random way, but they are unable to account for the complex organized design in organisms and are, correspondingly, usually peripheralized into those aspects that do not make a significant impact on the functional operation of the system (Tooby and Cosmides, 1990a, 1990b, 1992). Random walks do not systematically build intricate and improbably functional arrangements such as the visual system, the language faculty, face recognition programs, emotion recognition modules, food aversion circuits, cheater detection devices, or motor control systems, for the same reason that wind in a junkyard does not assemble airplanes and radar.

Brains Are Composed Primarily of Adaptive Problem-Solving Devices

In fact, natural selection is the only known cause of and explanation for complex functional design in organic systems. Hence, all naturally occurring functional organization in organisms should be ascribed to its operation, and hypotheses about function are likely to be correct only if they are the kinds of functionality that natural selection produces.

This leads to the most important point for cognitive neuroscientists to abstract from modern evolutionary biology: Although not everything in the designs of organisms is the product of selection, all complex functional organization is. Indeed, selection can only account for functionality of a very narrow kind: approximately, design features organized to promote the reproduction of an individual and his or her relatives in ancestral environments (Williams, 1966; Dawkins, 1986). Fortunately for the modern theory of evolution, the only naturally occurring complex functionality that ever has been documented in undomesticated plants, animals, or other organisms is functionality of just this kind, along with its derivatives and by-products.

This has several important implications for cognitive neuroscientists:

1. *Technical definition of function*. In explaining or exploring the reliably developing organization of a cognitive device, the *function* of a design refers solely to how it systematically caused its own propagation in ancestral environments. It does not validly refer to any intuitive or folk definitions of function such as "contributing to personal goals," "contributing to one's well-being," or "contributing to society." These other kinds of usefulness may or may not exist as side effects of a given evolved design, but they can play no role in explaining how such designs came into existence or why they have the organization that they do.

It is important to bear in mind that the evolutionary standard of functionality is entirely independent of any ordinary human standard of desirability, social value, morality, or health (Cosmides and Tooby, 1999).

2. *Adapted to the past*. The human brain, to the extent that it is organized to do anything functional at all, is organized to construct information, make decisions, and generate behavior that would have tended to promote inclusive fitness in the ancestral environments and behavioral contexts of Pleistocene hunter-gatherers and before. (The preagricultural world of hunter-gatherers is the appropriate ancestral context because natural selection operates far too slowly to have built complex information-processing adaptations to the post-hunter-gatherer world of the last few thousand years.)

3. *No evolved "reading modules*." The problems that our cognitive devices are designed to solve do not reflect the problems that our modern life experiences lead us to see as normal, such as reading, driving cars, working for large organizations, reading insurance forms, learning the oboe, or playing Go. Instead, they are the odd and seemingly esoteric problems that our hunter-gatherer ancestors encountered generation after generation over hominid evolution. These include such problems as foraging, kin recognition, "mind reading" (i.e., inferring beliefs, desires, and intentions from behavior), engaging in social exchange, avoiding incest, choosing mates, interpreting threats, recognizing emotions, caring for children, regulating immune function, and so on, as well as the already well-known problems involved in perception, language acquisition, and motor control.

4. *Side effects are personally important but scientifically misleading*. Although our architectures may be capable of performing tasks that are "functional" in the (nonbiological) sense that we may value them (e.g., weaving, playing piano), these are incidental side effects of selection for our Pleistocene competencies—just as a machine built to be a hair-dryer can, incidentally, dehydrate fruit or electrocute. But it will be difficult to make sense of our cognitive mechanisms if one attempts to interpret them as devices designed to perform functions that were not selectively important for our hunter-gatherer ancestors, or if one fails to consider the adaptive functions these abilities are side effects of.

5. *Adaptationism provides new techniques and principles*. Whenever one finds better-than-chance functional organization built into our cognitive or neural architecture, one is looking at adaptations—devices that acquired their distinctive organization from natural selection acting on our hunter-gatherer or more distant primate ancestors. Reciprocally, when one is searching for intelligible functional organization underlying a set of cognitive or neural phenomena, one

is far more likely to discover it by using an adaptationist framework for organizing observations because adaptive organization is the only kind of functional organization that is there to be found.

Because the reliably developing mechanisms (i.e., circuits, modules, functionally isolable units, mental organs, or computational devices) that cognitive neuroscientists study are evolved adaptations, all the biological principles that apply to adaptations apply to cognitive devices. This connects cognitive neuroscience and evolutionary biology in the most direct possible way. This conclusion should be a welcome one because it is the logical doorway through which a very extensive body of new expertise and principles can be made to apply to cognitive neuroscience, stringently constraining the range of valid hypotheses about the functions and structures of cognitive mechanisms. Because cognitive neuroscientists are usually studying adaptations and their effects, they can supplement their present research methods with carefully derived adaptationist analytic tools.

6. *Ruling out and ruling in*. Evolutionary biology gives specific and rigorous content to the concept of function, imposing strict rules on its use (Williams, 1966; Dawkins, 1982, 1986). This allows one to rule out certain hypotheses about the proposed function of a given cognitive mechanism. But the problem is not just that cognitive neuroscientists sometimes impute functions that they ought not to. An even larger problem is that many fail to impute functions that they ought to. For example, an otherwise excellent recent talk by a prominent cognitive neuroscientist began with the claim that one would not expect jealousy to be a "primary" emotion—that is, a universal, reliably developing part of the human neural architecture (in contrast to others, such as disgust or fear). Yet there is a large body of theory in evolutionary biology—sexual selection theory—that predicts that sexual jealousy will be widespread in species with substantial parental investment in offspring (particularly in males); behavioral ecologists have documented mate-guarding behavior (behavior designed to keep sexual competitors away from one's mate) in a wide variety of species, including various birds, fish, insects, and mammals (Krebs and Davies, 1997; Wilson and Daly, 1992); male sexual jealousy exists in every documented human culture (Daly et al., 1982; Wilson and Daly, 1992); it is the major cause of spousal homicides (Daly and Wilson, 1988), and in experimental settings, the design features of sexual jealousy have been shown to differ between the sexes in ways that reflect the different adaptive problems faced by ancestral men and women (Buss, 1994). From the standpoint of evolutionary biology and behavioral ecology, the hypothesis that sexual jealousy is a primary emotion—more specifically, the hypothesis that the human brain includes neurocognitive mechanisms whose function is to regulate the conditions under which sexual jealousy is expressed and what its cognitive and behavioral manifestations will be like—is virtually inescapable (for an evolutionary/cognitive approach to emotions, see Tooby and Cosmides, 1990a, 1990b). But if cognitive neuroscientists are not aware of this body of theory and evidence, they will not design experiments capable of revealing such mechanisms.

7. *Biological parsimony, not physics parsimony*. The standard of parsimony imported from physics, the traditional philosophy of science, or from habits of economical programming is inappropriate and misleading in biology, and

hence, in neuroscience and cognitive science, which study biological systems. The evolutionary process never starts with a clean work board, has no foresight, and incorporates new features solely on the basis of whether they lead to systematically enhanced propagation. Indeed, when one examines the brain, one sees an amazingly heterogeneous physical structure. A correct theory of evolved cognitive functions should be no less complex and heterogeneous than the evolved physical structure itself and should map on to the heterogeneous set of recurring adaptive tasks faced by hominid foragers over evolutionary time. Theories of engineered machinery involve theories of the subcomponents. One would not expect that a general, unified theory of robot or automotive mechanism could be accurate.

8. *Many cognitive adaptations*. Indeed, analyses of the adaptive problems humans and other animals must have regularly solved over evolutionary time suggest that the mind contains a far greater number of functional specializations than is traditionally supposed, even by cognitive scientists sympathetic to "modular" approaches. From an evolutionary perspective, the human cognitive architecture is far more likely to resemble a confederation of hundreds or thousands of functionally dedicated computers, designed to solve problems endemic to the Pleistocene, than it is to resemble a single general purpose computer equipped with a small number of domain-general procedures, such as association formation, categorization, or production rule formation (for discussion, see Cosmides and Tooby, 1987, 1994; Gallistel, 1990; Pinker, 1997; Sperber, 1994; Symons, 1987; Tooby and Cosmides, 1992).

9. *Cognitive descriptions are necessary*. Understanding the neural organization of the brain depends on understanding the functional organization of its computational relationships or cognitive devices. The brain originally came into existence and accumulated its particular set of design features only because these features functionally contributed to the organism's propagation. This contribution—that is, the evolutionary function of the brain—is obviously the adaptive regulation of behavior and physiology *on the basis of information* derived from the body and from the environment. The brain performs no significant mechanical, metabolic, or chemical service for the organism—its function is purely informational, computational, and regulatory in nature. Because the function of the brain is informational in nature, its precise functional organization can only be accurately described in a language that is capable of expressing its informational functions—that is, in cognitive terms, rather than in cellular, anatomical, or chemical terms. Cognitive investigations are not some soft, optional activity that goes on only until the "real" neural analysis can be performed. Instead, the mapping of the computational adaptations of the brain is an unavoidable and indispensable step in the neuroscience research enterprise. It must proceed in tandem with neural investigations and provides one of the primary frameworks necessary for organizing the body of neuroscience results.

The reason is straightforward. Natural selection retained neural structures on the basis of their ability to create adaptively organized relationships between information and behavior (e.g., the sight of a predator activates inference procedures that cause the organism to hide or flee) or between information and physiology (e.g., the sight of a predator increases the organism's heart rate, in preparation for flight). Thus, it is the information-processing structure of the

human psychological architecture that has been functionally organized by natural selection, and the neural structures and processes have been organized insofar as they physically realize this cognitive organization. Brains exist and have the structure that they do because of the computational requirements imposed by selection on our ancestors. The adaptive structure of our computational devices provides a skeleton around which a modern understanding of our neural architecture should be constructed.

Brain Architectures Consist of Adaptations, By-Products, and Random Effects

To understand the human (or any living species') computational or neural architecture is a problem in reverse engineering: We have working exemplars of the design in front of us, but we need to organize our observations of these exemplars into a systematic functional and causal description of the design. One can describe and decompose brains into properties according to any of an infinite set of alternative systems, and hence there are an indefinitely large number of cognitive and neural phenomena that could be defined and measured. However, describing and investigating the architecture in terms of its adaptations is a useful place to begin, because (1) the adaptations are the cause of the system's organization (the reason for the system's existence), (2) organisms, properly described, consist largely of collections of adaptations (evolved problem-solvers), (3) an adaptationist frame of reference allows cognitive neuroscientists to apply to their research problems the formidable array of knowledge that evolutionary biologists have accumulated about adaptations, (4) all of the complex functionally organized subsystems in the architecture are adaptations, and (5) such a frame of reference permits the construction of economical and principled models of the important features of the system, in which the wealth of varied phenomena fall into intelligible, functional, and predictable patterns. As Ernst Mayr put it, summarizing the historical record, "the adaptationist question, 'What is the function of a given structure or organ?' has been for centuries the basis for every advance in physiology" (Mayr, 1983, p. 32). It should prove no less productive for cognitive neuroscientists. Indeed, all of the inherited design features of organisms can be partitioned into three categories: (1) adaptations (often, although not always, complex); (2) the by-products or concomitants of adaptations; and (3) random effects. Chance and selection, the two components of the evolutionary process, explain different types of design properties in organisms, and all aspects of design must be attributed to one of these two forces. The conspicuously distinctive cumulative impacts of chance and selection allow the development of rigorous standards of evidence for recognizing and establishing the existence of adaptations and distinguishing them from the nonadaptive aspects of organisms caused by the nonselectionist mechanisms of evolutionary change (Williams, 1966, 1985; Pinker and Bloom, 1992; Symons, 1992; Thornhill, 1991; Tooby and Cosmides, 1990a, 1990b, 1992; Dawkins, 1986).

Design Evidence

Adaptations are systems of properties ("mechanisms") crafted by natural selection to solve the specific problems posed by the regularities of the physical, chemical, developmental, ecological, demographic, social, and informational

Table 29.1
The formal properties of an adaptation

An adaptation is:

1. A cross-generationally recurring set of characteristics of the phenotype
2. that is reliably manufactured over the developmental life history of the organism,
3. according to instructions contained in its genetic specification,
4. in interaction with stable and recurring features of the environment (i.e., it reliably develops normally when exposed to normal ontogenetic environments),
5. whose genetic basis became established and organized in the species (or population) over evolutionary time, because
6. the set of characteristics systematically interacted with stable and recurring features of the ancestral environment (the "adaptive problem"),
7. in a way that systematically promoted the propagation of the genetic basis of the set of characteristics better than the alternative designs existing in the population during the period of selection. This promotion virtually always takes place through enhancing the reproduction of the individual bearing the set of characteristics, or the reproduction of the relatives of that individual.

Adaptations. The most fundamental analytic tool for organizing observations about a species' functional architecture is the definition of an adaptation. To function, adaptations must evolve such that their causal properties rely on and exploit these stable and enduring statistical structural regularities in the world, and in other parts of the organism. Things worth noticing include the fact that an adaptation (such as teeth or breasts) can develop at any time during the life cycle, and need not be present at birth; an adaptation can express itself differently in different environments (e.g., speaks English, speaks Tagalog); an adaptation is not just any individually beneficial trait, but one built over evolutionary time and expressed in many individuals; an adaptation may not be producing functional outcomes currently (e.g., agoraphobia), but only needed to function well in ancestral environments; finally, an adaptation (like every other aspect of the phenotype) is the product of gene–environment interaction. Unlike many other phenotypic properties, however, it is the result of the interaction of the species-standard set of genes with those aspects of the environment that were present and relevant during the species' evolution. For a more extensive definition of the concept of adaptation, see Tooby and Cosmides, 1990b, 1992.

environments encountered by ancestral populations during the course of a species' or population's evolution (table 29. 1). Adaptations are recognizable by "evidence of special design" (Williams, 1966)—that is, by recognizing certain features of the evolved species-typical design of an organism "as components of some special problem-solving machinery" (Williams, 1985, p. 1). Moreover, they are so well organized and such good engineering solutions to adaptive problems that a chance coordination between problem and solution is effectively ruled out as a counter-hypothesis. Standards for recognizing special design include whether the problem solved by the structure is an evolutionarily long-standing adaptive problem, and such factors as economy, efficiency, complexity, precision, specialization, and reliability, which, like a key fitting a lock, render the design too good a solution to a defined adaptive problem to be coincidence (Williams, 1966). Like most other methods of empirical hypothesis testing, the demonstration that something is an adaptation is always, at core, a probability assessment concerning how likely a set of events is to have arisen by chance alone. Such assessments are made by investigating whether there is a highly nonrandom coordination between the recurring properties of the phenotype and the structured properties of the adaptive problem, in a way that meshed to promote fitness (genetic propagation) in ancestral environments

(Tooby and Cosmides, 1990b, 1992). For example, the lens, pupil, iris, retina, visual cortex, and other parts of the eye are too well coordinated, both with each other and with features of the world, such as the properties of light, optics, geometry, and the reflectant properties of surfaces, to have co-occurred by chance. In short, like the functional aspects of any other engineered system, they are recognizable as adaptations for analyzing scenes from reflected light by their organized and functional relationships to the rest of the design and to the structure of the world.

In contrast, concomitants or by-products of adaptations are those properties of the phenotype that do not contribute to functional design per se, but that happen to be coupled to properties that are. Consequently, they were dragged along into the species-typical architecture because of selection for the functional design features to which they are linked. For example, bones are adaptations, but the fact that they are white is an incidental by-product. Bones were selected to include calcium because it conferred hardness and rigidity to the structure (and was dietarily available), and it simply happens that alkaline earth metals appear white in many compounds, including the insoluble calcium salts that are a constituent of bone. From the point of view of functional design, by-products are the result of "chance," in the sense that the process that led to their incorporation into the design was blind to their consequences (assuming that they were not negative). Accordingly, such by-products are distinguishable from adaptations by the fact that they are not complexly arranged to have improbably functional consequences (e.g., the whiteness of bone does nothing for the vertebrae).

In general, by-products will be far less informative as a focus of study than adaptations because they are consequences and not causes of the organization of the system (and hence are functionally arbitrary, unregulated, and may, for example, vary capriciously between individuals). Unfortunately, unless researchers actively seek to study organisms in terms of their adaptations, they usually end up measuring and investigating arbitrary and random admixtures of functional and functionless aspects of organisms, a situation that hampers the discovery of the underlying organization of the biological system. We do not yet, for example, even know which exact aspects of the neuron are relevant to its function and which are by-products, so many computational neuroscientists may be using a model of the neuron that is wildly inaccurate.

Finally, entropic effects of many types are always acting to introduce disorder into the design of organisms. Traits introduced by accident or by evolutionary random walks are recognizable by the lack of coordination that they produce within the architecture or between the architecture and the environment, as well as by the fact that they frequently cause uncalibrated variation between individuals. Examples of such entropic processes include genetic mutation, recent change in ancestrally stable environmental features, and developmentally anomalous circumstances.

How Well-Engineered Are Adaptations?

The design of our cognitive and neural mechanisms should only reflect the structure of the adaptive problems that our ancestors faced to the extent that natural selection is an effective process. Is it one? How well or poorly engi-

neered are adaptations? Some researchers have argued that evolution primarily produces inept designs, because selection does not produce perfect optimality (Gould and Lewontin, 1979). In fact, evolutionary biologists since Darwin have been well aware that selection does not produce perfect designs (Darwin, 1859; Williams, 1966; Dawkins, 1976, 1982, 1986; for a recent convert from the position that organisms are optimally designed to the more traditional adaptationist position, see Lewontin, 1967, 1979; see Dawkins, 1982, for an extensive discussion of the many processes that prevent selection from reaching perfect optimality). Still, because natural selection is a hill-climbing process that tends to choose the best of the variant designs that actually appear, and because of the immense numbers of alternatives that appear over the vast expanse of evolutionary time, natural selection tends to cause the accumulation of very well-engineered functional designs.

Empirical confirmation can be gained by comparing how well evolved devices and human engineered devices perform on evolutionarily recurrent adaptive problems (as opposed to arbitrary, artificial modern tasks, such as chess). For example, the claim that language competence is a simple and poorly engineered adaptation cannot be taken seriously, given the total amount of time, engineering, and genius that has gone into the still unsuccessful effort to produce artificial systems that can remotely approach—let alone equal—human speech perception, comprehension, acquisition, and production (Pinker and Bloom, 1992).

Even more strikingly, the visual system is composed of collections of cognitive adaptations that are well-engineered products of the evolutionary process, and although they may not be "perfect" or "optimal"—however these somewhat vague concepts may be interpreted—they are far better at vision than any human-engineered system yet developed.

Wherever the standard of biological functionality can be clearly defined—semantic induction, object recognition, color constancy, echolocation, relevant problem-solving generalization, chemical recognition (olfaction), mimicry, scene analysis, chemical synthesis—evolved adaptations are at least as good as and usually strikingly better than human engineered systems, in those rare situations in which humans can build systems that can accomplish them at all. It seems reasonable to insist that before a system is criticized as being poorly designed, the critic ought to be able to construct a better alternative—a requirement, it need hardly be pointed out, that has never been met by anyone who has argued that adaptations are poorly designed. Thus, although adaptations are certainly suboptimal in some ultimate sense, it is an empirically demonstrable fact that the short-run constraints on selective optimization do not prevent the emergence of superlatively organized computational adaptations in brains. Indeed, aside from the exotic nature of the problems that the brain was designed to solve, it is exactly this sheer functional intricacy that makes our architecture so difficult to reverse-engineer and to understand.

Cognitive Adaptations Reflect the Structure of the Adaptive Problem and the Ancestral World

Four lessons emerge from the study of natural competences, such as vision and language: (1) most adaptive information-processing problems are complex; (2)

the evolved solution to these problems is usually machinery that is well engineered for the task; (3) this machinery is usually specialized to fit the particular nature of the problem; and (4) its evolved design often embodies substantial and contentful "innate knowledge" about problem-relevant aspects of the world.

Well-studied adaptations overwhelmingly achieve their functional outcomes because they display an intricately engineered coordination between their specialized design features and the detailed structure of the task and task environment. Like a code that has been torn in two and given to separate couriers, the two halves (the structure of the mechanism and the structure of the task) must be put together to be understood. To function, adaptations evolve such that their causal properties rely on and exploit these stable and enduring statistical and structural regularities in the world. Thus, to map the structures of our cognitive devices, we need to understand the structures of the problems that they solve and the problem-relevant parts of the hunter-gatherer world. If studying face recognition mechanisms, one must study the recurrent structure of faces. If studying social cognition, one must study the recurrent structure of hunter-gatherer social life. For vision, the problems are not so very different for a modern scientist and a Pleistocene hunter-gatherer, so the folk notions of function that perception researchers use are not a problem. But the more one strays from low-level perception, the more one needs to know about human behavioral ecology and the structure of the ancestral world.

Experimenting with Ancestrally Valid Tasks and Stimuli

Although bringing cognitive neuroscience current with modern evolutionary biology offers many new research tools (Preuss, 1995), we have out of necessity limited discussion to only one: an evolutionary functionalist research strategy (see Tooby and Cosmides, 1992, for a description; for examples, see chapters in Barkow et al., 1992; Daly and Wilson, 1995; Gaulin, 1995). The adoption of such an approach will modify research practice in many ways. Perhaps most significantly, researchers will no longer have to operate purely by intuition or guesswork to know which kinds of tasks and stimuli to expose subjects to. Using knowledge from evolutionary biology, behavioral ecology, animal behavior, and hunter-gatherer studies, they can construct ancestrally or adaptively valid stimuli and tasks. These are stimuli that would have had adaptive significance in ancestral environments, and tasks that resemble (at least in some ways) the adaptive problems that our ancestors would have been selected to be able to solve.

The present widespread practice of using arbitrary stimuli of no adaptive significance (e.g., lists of random words, colored geometric shapes) or abstract experimental tasks of unknown relevance to Pleistocene life has sharply limited what researchers have observed and can observe about our evolved computational devices. This is because the adaptive specializations that are expected to constitute the majority of our neural architecture are designed to remain dormant until triggered by cues of the adaptively significant situations that they were designed to handle. The Wundtian and British Empiricist methodological assumption that complex stimuli, behaviors, representations, and competences are compounded out of simple ones has been empirically falsified in scores

of cases (see, e.g., Gallistel, 1990), and so, restricting experimentation to such stimuli and tasks simply restricts what researchers can find to a highly impoverished and unrepresentative set of phenomena. In contrast, experimenters who use more biologically meaningful stimuli have had far better luck, as the collapse of behaviorism and its replacement by modern behavioral ecology have shown in the study of animal behavior. To take one example of its applicability to humans, effective mechanisms for Bayesian inference—undetected by 20 years of previous research using "modern" tasks and data formats—were activated by exposing subjects to information formatted in a way that hunter-gatherers would have encountered it (Brase et al., 1998; Cosmides and Tooby, 1996; Gigerenzer and Hoffrage, 1995). Equally, when subjects were given ancestrally valid social inference tasks (cheater detection, threat interpretation), previously unobserved adaptive reasoning specializations were activated, guiding subjects to act in accordance with evolutionarily predicted but otherwise odd patterns (Cosmides, 1989; Cosmides and Tooby, 1992).

Everyone accepts that one cannot study human language specializations by exposing subjects to meaningless sounds: the acoustic stimuli must contain the subtle, precise, high level relationships that make sound language. Similarly, to move on to the study of other complex cognitive devices, subjects should be exposed to stimuli that contain the subtle, ancestrally valid relationships relevant to the diverse functions of these devices. In such an expanded research program, experimental stimuli and tasks would involve constituents such as faces, smiles, disgust expressions, foods, the depiction of socially significant situations, sexual attractiveness, habitat quality cues, animals, navigational problems, cues of kinship, rage displays, cues of contagion, motivational cues, distressed children, species-typical "body language," rigid object mechanics, plants, predators, and other functional elements that would have been part of ancestral hunter-gatherer life. Investigations would look for functional subsystems that not only deal with such low-level and broadly functional competences as perception, attention, memory, and motor control, but also with higher-level ancestrally valid competences as well—mechanisms such as eye direction detectors (Baron-Cohen, 1994), face recognizers (e.g., Johnson and Morton, 1991), food memory subsystems (e.g., Hart et al., 1985; Caramazza and Shelton, 1998), person-specific memory, child care motivators (Daly and Wilson, 1995), and sexual jealousy modules.

Although these proposals to look for scores of content-sensitive circuits and domain-specific specializations will strike many as bizarre and even preposterous, they are well grounded in modern biology. We believe that in a decade or so they will look tame. If cognitive neuroscience is anything like investigations in domain-specific cognitive psychology (Hirschfeld and Gelman, 1994) and in modern animal behavior, researchers will be rewarded with the materialization of a rich array of functionally patterned phenomena that have not been observed so far because the mechanisms were never activated in the laboratory by exposure to ecologically appropriate stimuli. Although presently, the functions of most brain structures are largely unknown, pursuing such research directions may begin to populate the empty regions of our maps of the brain with circuit diagrams of discrete, functionally intelligible computational devices.

In short, because theories and principled systems of knowledge can function as organs of perception, the incorporation of a modern evolutionary framework into cognitive neuroscience may allow the community to detect ordered relationships in phenomena that otherwise seem too complex to be understood.

Conclusion

The aforementioned points indicate why cognitive neuroscience is pivotal to the progress of the brain sciences. There are an astronomical number of physical interactions and relationships in the brain, and blind empiricism rapidly drowns itself among the deluge of manic and enigmatic measurements. Through blind empiricism, one can equally drown at the cognitive level in a sea of irrelevant things that our computational devices can generate, from writing theology or dancing the mazurka to calling for the restoration of the Plantagenets to the throne of France. However, evolutionary biology, behavioral ecology, and hunter-gatherer studies can be used to identify and supply descriptions of the recurrent adaptive problems humans faced during their evolution. Supplemented with this knowledge, cognitive research techniques can abstract out of the welter of human cognitive performance a series of maps of the functional information-processing relationships that constitute our computational devices and that evolved to solve this particular set of problems: our cognitive architecture. These computational maps can then help us abstract out of the ocean of physical relationships in the brain that exact and minute subset that implements those information-processing relationships because it is only these relationships that explain the existence and functional organization of the system. The immense number of other physical relationships in the brain are incidental by-products of those narrow aspects that implement the functional computational architecture. Consequently, an adaptationist inventory and functional mapping of our cognitive devices can provide the essential theoretical guidance for neuroscientists that will allow them to home in on these narrow but meaningful aspects of neural organization and to distinguish them from the sea of irrelevant neural phenomena.

Acknowledgments

The authors gratefully acknowledge the financial support of the James S. McDonnell Foundation, the National Science Foundation (NSF grant BNS9157-449 to John Tooby), and a Research Across Disciplines grant (Evolution and the Social Mind) from the UCSB Office of Research.

Note

1. The genes underlying complex adaptations cannot vary substantially between individuals because if they did, the obligatory genetic shuffling that takes place during sexual reproduction would break apart the complex adaptations that had existed in the parents when these are recombined in the offspring generation. All the genetic subcomponents necessary to build the complex adaptation rarely would reappear together in the same individual if they were not being supplied reliably by both parents in all matings (for a discussion of the genetics of sexual recombination, species-typical adaptive design, and individual differences, see Tooby, 1982; Tooby and Cosmides, 1990b).

References

Barkow, J., L. Cosmides, and J. Tooby, eds., 1992. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

Baron-Cohen, S., 1994. The eye-direction detector: A case for evolutionary psychology. In *Joint-Attention: Its Origins and Role in Development*, C. Moore and P. Dunham, eds. Hillsdale, N.J.: Erlbaum.

Brase, G., L. Cosmides, and J. Tooby, 1998. Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. *J. Exp. Psychol. Gen*. 127:3–21.

Buss, D., 1994. *The Evolution of Desire*. New York: Basic Books.

Caramazza, A., and J. Shelton, 1998. Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *J. Cogn. Neurosci*. 10:1–34.

Cosmides, L., 1989. The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31:187–276.

Cosmides, L., and J. Tooby, 1981. Cytoplasmic inheritance and intragenomic conflict. *J. Theor. Biol*. 89:83–129.

Cosmides, L., and J. Tooby, 1987. From evolution to behavior: Evolutionary psychology as the missing link. In *The Latest on the Best: Essays on Evolution and Optimality*, J. Dupre, ed. Cambridge, Mass.: MIT Press, pp. 277–306.

Cosmides, L., and J. Tooby, 1992. Cognitive adaptations for social exchange. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, J. Barkow, L. Cosmides, and J. Tooby, eds. New York: Oxford University Press, pp. 163–228.

Cosmides, L., and J. Tooby, 1994. Beyond intuition and instinct blindness: The case for an evolutionarily rigorous cognitive science. *Cognition* 50:41–77.

Cosmides, L., and J. Tooby, 1996. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition* 58:1–73.

Cosmides, L., and J. Tooby, 1999. Toward an evolutionary taxanomy of treatable conditions. *J. Abnorm. Psychol.* 108:435–464.

Daly, M., and M. Wilson, 1984. *Sex, Evolution and Behavior*, Second Edition. Boston: Willard Grant.

Daly, M., and M. Wilson, 1988. *Homicide*. New York: Aldine.

Daly, M., and M. Wilson, 1995. Discriminative parental solicitude and the relevance of evolutionary models to the analysis of motivational systems. In *The Cognitive Neurosciences*, M. S. Gazzaniga, ed. Cambridge, Mass.: MIT Press, pp. 1269–1286.

Daly, M., M. Wilson, and S. J. Weghorst, 1982. Male sexual jealousy. *Ethol. Sociobiol*. 3:11–27.

Darwin, C., 1859. *On the Origin of Species*. London: Murray. New edition: Cambridge, Mass.: Harvard University Press.

Dawkins, R., 1976. *The Selfish Gene*. New York: Oxford University Press.

Dawkins, R., 1982. *The Estended Phenotype*. San Francisco: W. H. Freeman.

Dawkins, R., 1986. *The Blind Watchmaker*. New York: Norton.

Gallistel, C. R., 1990. *The Organization of Learning*. Cambridge, Mass.: MIT Press.

Gaulin, S., 1995. Does evolutionary theory predict sex differences in the brain? In *The Cognitive Neurosciences*, M. S. Gazzaniga, ed. Cambridge, Mass.: MIT Press, pp. 1211–1225.

Gigerenzer, G., and U. Hoffrage, 1995. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychol. Rev*. 102:684–704.

Gould, S. J., and R. C. Lewontin, 1979. The spandrels of San Marco and the Panglossian program: A critique of the adaptationist programme. *Proc. R. Soc. Lond*. 205:281–288.

Haig, D., 1993. Genetic conflicts in human pregnancy. *Q. Rev. Biol*. 68:495–532.

Hamilton, W. D., 1964. The genetical evolution of social behavior. *J. Theor. Biol*. 7:1–52.

Hart, J. Jr, R. S. Berndt, and A. Caramazza, 1985. Category-specific naming deficit following cerebral infarction. *Nature* 316:439–440.

Hirschfeld, L., and S. Gelman, eds. 1994. *Mapping the Mind: Domain Specificity in Cognition and Culture*. New York: Cambridge University Press.

Johnson, M., and J. Morton, 1991. *Biology and Cognitive Development: The Case of Face Recognition*. Oxford: Blackwell.

Krebs, J. R., and N. B. Davies, 1997. *Behavioural Ecology: An Evolutionary Approach*, 4th edition. London: Blackwell Science.

Lee, R. B., and I. DeVore, 1968. *Man the Hunter*. Chicago: Aldine.

Lewontin, R., 1967. Spoken remark in *Mathematical Challenges to the Neo-Darwinian Interpretation of Evolution*, P. Moorhead and M. Kaplan, eds. *Wistar Institute Symposium Monograph* 5:79.

Lewontin, R., 1979. Sociobiology as an adaptationist program. *Behav. Sci.* 24:5–14.

Marr, D., 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.

Mayr, E., 1983. How to carry out the adaptationist program. *Am. Naturalist* 121:324–334.

Pinker, S., 1994. *The Language Instinct*. New York: Morrow.

Pinker, S., 1997. *How the Mind Works*. New York: Norton.

Pinker, S., and P. Bloom, 1992. Natural language and natural selection. Reprinted in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, J. Barkow, L. Cosmides, and J. Tooby, eds. New York: Oxford University Press, pp. 451–493.

Preuss, T., 1995. The argument from animals to humans in cognitive neuroscience. In *The Cognitive Neurosciences*, M. S. Gazzaniga, ed. Cambridge, Mass.: MIT Press, pp. 1227–1241.

Shepard, R. N., 1984. Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychol. Rev.* 91:417–447.

Shepard, R. N., 1987a. Evolution of a mesh between principles of the mind and regularities of the world. In *The Latest on the Best: Essays on Evolution and Optimality*, J. Dupre, ed. Cambridge, Mass.: MIT Press, pp. 251–275.

Shepard, R. N., 1987b. Towards a universal law of generalization for psychological science. *Science* 237:1317–1323.

Sperber, D., 1994. The modularity of thought and the epidemiology of representations. *In Mapping the Mind: Domain Specificity in Cognition and Culture*, L. Hirschfeld and S. Gelman, eds. New York: Cambridge University Press, pp. 39–67.

Symons, D., 1987. If we're all Darwinians, what's the fuss about? In *Sociobiology and Psychology*, C. B. Crawford, M. F. Smith, and D. L. Krebs, eds. Hillsdale, N.J.: Erlbaum, pp. 121–146.

Symons, D., 1992. On the use and misuse of Darwinism in the study of human behavior. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, J. Barkow, L. Cosmides, and J. Tooby, eds. New York: Oxford University Press, pp. 137–159.

Thornhill, R., 1991. The study of adaptation. In *Interpretation and Explanation in the Study of Behavior*, M. Bekoff and D. Jamieson, eds. Boulder, Colo.: Westview Press.

Tooby, J., 1982. Pathogens, polymorphism, and the evolution of sex. *J. Theor. Biol.* 97:557–576.

Tooby, J., and L. Cosmides, 1990a. The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethol. Sociobiol.* 11:375–424.

Tooby, J., and L. Cosmides, 1990b. On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *J. Pers.* 58:17–67.

Tooby, J., and L. Cosmides, 1992. The psychological foundations of culture. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, J. Barkow, L. Cosmides, and J. Tooby, eds. New York: Oxford University Press, pp. 19–136.

Williams, G. C., 1966. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton, N.J.: Princeton University Press.

Williams, G. C., 1985. A defense of reductionism in evolutionary biology. *Oxford Surv. Biol.* 2:1–27.

Williams, G. C., and D. C. Williams, 1957. Natural selection of individually harmful social adaptations among sibs with special reference to social insects. *Evolution* 17:249–253.

Wilson, M., and M. Daly, 1992. The man who mistook his wife for a chattel. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, J. Barkow, L. Cosmides, and J. Tooby, eds. New York: Oxford University Press, pp. 289–322.

PART XIV

Language 1—Language Acquisition

PART XV

Language 2—Language and Thought

Chapter 31

Languages and Logic

Benjamin L. Whorf

In English, the sentences 'I pull the branch aside' and 'I have an extra toe on my foot' have little similarity. Leaving out the subject pronoun and the sign of the present tense, which are common features from requirements of English syntax, we may say that no similarity exists. Common, and even scientific, parlance would say that the sentences are unlike because they are talking about things which are intrinsically unlike. So Mr. Everyman, the natural logician, would be inclined to argue. Formal logic of an older type would perhaps agree with him.

If, moreover, we appeal to an impartial scientific English-speaking observer, asking him to make direct observations upon cases of the two phenomena to see if they may not have some element of similarity which we have overlooked, he will be more than likely to confirm the dicta of Mr. Everyman and the logician. The observer whom we have asked to make the test may not see quite eye to eye with the old-school logician and would not be disappointed to find him wrong. Still he is compelled sadly to confess failure. "I wish I could oblige you," he says, "but try as I may, I cannot detect any similarity between these phenomena."

By this time our stubborn streak is aroused; we wonder if a being from Mars would also see no resemblance. But now a linguist points out that it is not necessary to go as far as Mars. We have not yet scouted around this earth to see if its many languages all classify these phenomena as disparately as our speech does. We find that in Shawnee these two statements are, respectively, *ni-l'θawa-'ko-n-a* and *ni-l'θawa-'ko-θite* (the *θ* here denotes *th* as in 'thin' and the apostrophe denotes a breath-catch). The sentences are closely similar; in fact, they differ only at the tail end. In Shawnee, moreover, the beginning of a construction is generally the important and emphatic part. Both sentences start with *ni-*('I'), which is a mere prefix. Then comes the really important key word, *l'θawa*, a common Shawnee term, denoting a forked outline, like figure 31.1, no. 1. The next element, *-'ko*, we cannot be sure of, but it agrees in form with a variant of the suffix *-a'kw* or *-a'ko*, denoting tree, bush, tree part, branch, or anything of that general shape. In the first sentence, *-n-* means 'by hand action' and may be either a cansation of the basic condition (forked outline) manually, an increase of it, or both. The final *-a* means that the subject ('I') does this action to an appropriate object. Hence the first sentence means 'I pull it (something like branch of tree) more open or apart where it forks.' In the other sentence, the suffix *-θite* means 'pertaining to the toes,' and the absence of further suffixes

From *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, ed. J. B. Carroll (Cambridge, MA: MIT Press, 1956), 233–245. Reprinted with permission.

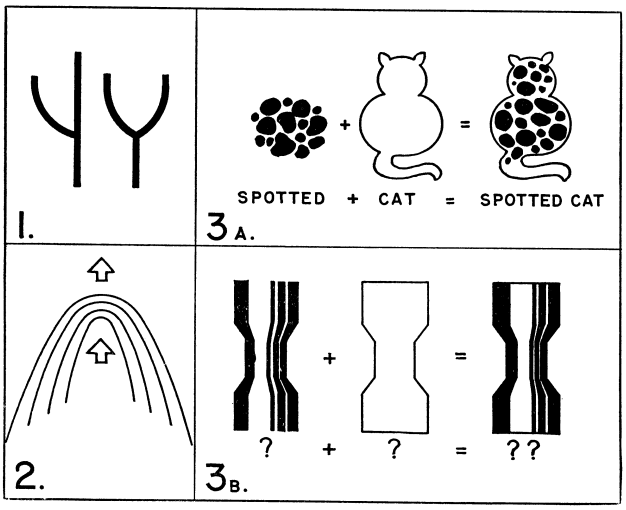


Figure 31.1
Suggested above are certain linguistic concepts which, as explained in the text, are not easily definable.

means that the subject manifests the condition in his own person. Therefore the sentence can mean only 'I have an extra toe forking out like a branch from a normal toe.'

Shawnee logicians and observers would class the two phenomena as intrinsically similar. Our own observer, to whom we tell all this, focuses his instruments again upon the two phenomena and to his joy sees at once a manifest resemblance. Figure 31.2 illustrates a similar situation: 'I push his head back' and 'I drop it in water and it floats,' though very dissimilar sentences in English, are similar in Shawnee. The point of view of linguistic relativity changes Mr. Everyman's dictum: Instead of saying, "Sentences are unlike because they tell about unlike facts," he now reasons: "Facts are unlike to speakers whose language background provides for unlike formulation of them."

Conversely, the English sentences, 'The boat is grounded on the beach' and 'The boat is manned by picked men,' seem to us to be rather similar. Each is about a boat; each tells the relation of the boat to other objects–or that's *our* story. The linguist would point out the parallelism in grammatical pattern thus: "The boat is *x*ed preposition *y*." The logician might turn the linguist's analysis into "*A* is in the state *x* in relation to *y*," and then perhaps into $fA = xRy$. Such symbolic methods lead to fruitful techniques of rational ordering, stimulate our thinking, and bring valuable insight. Yet we should realize that the similarities and contrasts in the original sentences, subsumed under the foregoing formula, are dependent on the choice of mother tongue and that the properties of the tongue are eventually reflected as peculiarities of structure in the fabric of logic or mathematics which we rear.

In the Nootka language of Vancouver Island, the first "boat" statement is *tlih-is-ma*; the second, *lash-tskwiq-ista-ma*. The first is thus I-II-*ma*; the second, III-IV-V-*ma*; and they are quite unlike, for the final *-ma* is only the sign of the third-

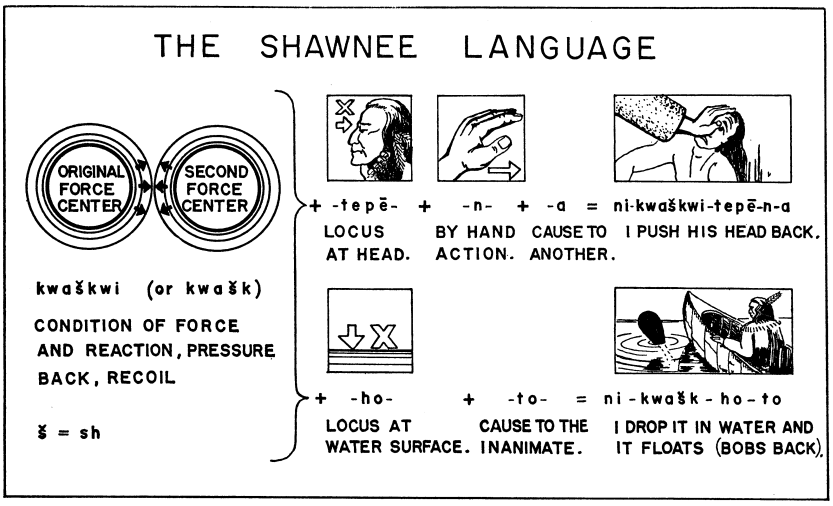


Figure 31.2
The English sentences 'I push his head back' and 'I drop it in water and it floats' are unlike. But in Shawnee the corresponding statements are closely similar, emphasizing the fact that analysis of nature and classification of events as like or in the same category (logic) are governed by grammar.

person indicative. Neither sentence contains any unit of meaning akin to our word 'boat' or even 'canoe.' Part I, in the first sentence, means 'moving pointwise,' or moving in a way like the suggestion of the outline in figure 31.1, no. 2; hence 'traveling in or as a canoe,' or an event like one position of such motion. It is not a name for what we should call a "thing," but is more like a vector in physics. Part II means 'on the beach'; hence I-II-*ma* means 'it is on the beach pointwise as an event of canoe motion,' and would normally refer to a boat that has come to land. In the other sentence, part III means 'select, pick,' and IV means 'remainder, result,' so that III-IV means 'selected.' Part V means 'in a canoe (boat) as crew.' The whole, III-IV-V-*ma*, means either 'they are in the boat as a crew of picked men' or 'the boat has a crew of picked men.' It means that the whole event involving picked ones and boat's crew is in process.

As a hang-over from my education in chemical engineering, I relish an occasional chemical simile. Perhaps readers will catch what I mean when I say that the way the constituents are put together in these sentences of Shawnee and Nootka suggests a chemical compound, whereas their combination in English is more like a mechanical mixture. A mixture, like the mountaineer's potlicker, can be assembled out of almost anything and does not make any sweeping transformation of the overt appearance of the material. A chemical compound, on the other hand, can be put together only out of mutually suited ingredients, and the result may be not merely soup but a crop of crystals or a cloud of smoke. Likewise the typical Shawnee or Nootka combinations appear to work with a vocabulary of terms chosen with a view not so much to the utility of their immediate references as to the ability of the terms to combine suggestively with each other in manifold ways that elicit novel and useful images. This principle of terminology and way of analyzing events would seem to be unknown to the tongues with which we are familiar.

It is the analysis of nature down to a basic vocabulary capable of this sort of evocative recombination which is most distinctive of polysynthetic languages, like Nootka and Shawnee. Their characteristic quality is not, as some linguists have thought, a matter of the tightness or indissolubility of the combinations. The Shawnee term *l'θawa* could probably be said alone but would then mean 'it (or something) is forked,' a statement which gives little hint of the novel meanings that arise out of its combinations—at least to our minds or our type of logic. Shawnee and Nootka do not use the chemical type of synthesis exclusively. They make large use of a more external kind of syntax, which, however, has no basic structural priority. Even our own Indo-European tongues are not wholly devoid of the chemical method, but they seldom make sentences by it, afford little inkling of its possibilities, and give structural priority to another method. It was quite natural, then, that Aristotle should found our traditional logic wholly on this other method.

Let me make another analogy, not with chemistry but with art—art of the pictorial sort. We look at a good still-life painting and seem to see a lustrous porcelain bowl and a downy peach. Yet an analysis that screened out the totality of the picture—as if we were to go over it carefully, looking through a hole cut in a card—would reveal only oddly shaped patches of paint and would not evoke the bowl and fruit. The synthesis presented by the painting is perhaps akin to the chemical type of syntax, and it may point to psychological fundamentals that enter into both art and language. Now the mechanical method in art and language might be typified by no. 3*A* in figure 31.1. The first element, a field of spots, corresponds to the adjective 'spotted,' the second corresponds to the noun 'cat.' By putting them together, we get 'spotted cat.' Contrast the technique in figure 31.1, no. 3*B*. Here the figure corresponding to 'cat' has only vague meaning by itself—"chevron-like," we might say—while the first element is even vaguer. But, combined, these evoke a cylindrical object, like a shaft casting.

The thing common to both techniques is a systematic synthetic use of pattern, and this is also common to all language techniques. I have put question marks below the elements in figure 31.1, no. 3*B*, to point out the difficulty of a parallel in English speech and the fact that the method probably has no standing in traditional logic. Yet examination of other languages and the possibility of new types of logic that has been advanced by modern logicians themselves suggest that this matter may be significant for modern science. New types of logic may help us eventually to understand how it is that electrons, the velocity of light, and other components of the subject matter of physics appear to behave illogically, or that phenomena which flout the sturdy common sense of yesteryear can nevertheless be true. Modern thinkers have long since pointed out that the so-called mechanistic way of thinking has come to an impasse before the great frontier problems of science. To rid ourselves of this way of thinking is exceedingly difficult when we have no linguistic experience of any other and when even our most advanced logicians and mathematicians do not provide any other—and obviously they cannot without the linguistic experience. For the mechanistic way of thinking is perhaps just a type of syntax natural to Mr. Everyman's daily use of the western Indo-European languages, rigidified and intensified by Aristotle and the latter's medieval and modern followers.

As I said in an article, "Science and linguistics," in the *Review* for April 1940, the effortlessness of speech and the subconscious way we picked up that activity in early childhood lead us to regard talking and thinking as wholly straightforward and transparent. We naturally feel that they embody self-evident laws of thought, the same for all men. We know all the answers! But, when scrutinized, they become dusty answers. We use speech for reaching agreements about subject matter: I say, "Please shut the door," and my hearer and I agree that 'the door' refers to a certain part of our environment and that I want a certain result produced. Our explanations of how we reached this understanding, though quite satisfactory on the everyday social plane, are merely more agreements (statements) about the same subject matter (door, and so on), more and more amplified by statements about the social and personal needs that impel us to communicate. There are here no laws of thought. Yet the structural regularities of our sentences enable us to sense that laws are *somewhere* in the background. Clearly, explanations of understanding such as "And so I ups and says to him, says I; see here, why don't you . . . !" evade the true process by which 'he' and 'I' are in communication. Likewise psychological-social descriptions of the social and emotional needs that impel people to communicate with their fellows tend to be learned versions of the same method and, while interesting, still evade the question. In similar case is evasion of the question by skipping from the speech sentence, via physiology and "stimuli," to the social situation.

The *why* of understanding may remain for a long time mysterious; but the *how* or logic of understanding—its background of laws or regularities—is discoverable. It is the grammatical background of our mother tongue, which includes not only our way of constructing propositions but the way we dissect nature and break up the flux of experience into objects and entities to construct propositions about. This fact is important for science, because it means that science *can* have a rational or logical basis even though it be a relativistic one and not Mr. Everyman's natural logic. Although it may vary with each tongue, and a planetary mapping of the dimensions of such variation may be necessitated, it is, nevertheless, a basis of logic with discoverable laws. Science is not compelled to see its thinking and reasoning procedures turned into processes merely subservient to social adjustments and emotional drives.

Moreover, the tremendous importance of language cannot, in my opinion, be taken to mean necessarily that nothing is back of it of the nature of what has traditionally been called "mind." My own studies suggest, to me, that language, for all its kingly role, is in some sense a superficial embroidery upon deeper processes of consciousness, which are necessary before any communication, signaling, or symbolism whatsoever can occur, and which also can, at a pinch, effect communication (though not true *agreement*) without language's and without symbolism's aid. I mean "superficial" in the sense that all processes of chemistry, for example, can be said to be superficial upon the deeper layer of physical existence, which we know variously as intra-atomic, electronic, or subelectronic. No one would take this statement to mean that chemistry is *unimportant*—indeed the whole point is that the more superficial can mean the more important, in a definite operative sense. It may even be in the cards that there is no such thing as "Language" (with a capital *L*) at all! The

statement that "thinking is a matter of *language*" is an incorrect generalization of the more nearly correct idea that "thinking is a matter of different tongues." The different tongues are the real phenomena and may generalize down not to any such universal as "Language," but to something better—called "sub-linguistic" or "superlinguistic"—and *not altogether* unlike, even if much unlike, what we now call "mental." This generalization would not diminish, but would rather increase, the importance of intertongue study for investigation of this realm of truth.

Botanists and zoologists, in order to understand the world of living species, found it necessary to describe the species in every part of the globe and to add a time perspective by including the fossils. Then they found it necessary to compare and contrast the species, to work out families and classes, evolutionary descent, morphology, and taxonomy. In linguistic science a similar attempt is under way. The far-off event toward which this attempt moves is a new technology of language and thought. Much progress has been made in classifying the languages of earth into genetic families, each having descent from a single precursor, and in tracing such developments through time. The result is called "comparative linguistics." Of even greater importance for the future technology of thought is what might be called "contrastive linguistics." This plots the outstanding differences among tongues—in grammar, logic, and general analysis of experience.

As I said in the April 1940 *Review*, segmentation of nature is an aspect of grammar—one as yet little studied by grammarians. We cut up and organize the spread and flow of events as we do, largely because, through our mother tongue, we are parties to an agreement to do so, not because nature itself is segmented in exactly that way for all to see. Languages differ not only in how they build their sentences but also in how they break down nature to secure the elements to put in those sentences. This breakdown gives units of the lexicon. "Word" is not a very good "word" for them; "lexeme" has been suggested, and "term" will do for the present. By these more or less distinct terms we ascribe a semifictitious isolation to parts of experience. English-terms, like 'sky, hill, swamp,' persuade us to regard some elusive aspect of nature's endless variety as a distinct *thing*, almost like a table or chair. Thus English and similar tongues lead us to think of the universe as a collection of rather distinct objects and events corresponding to words. Indeed this is the implicit picture of classical physics and astronomy—that the universe is essentially a collection of detached objects of different sizes.

The examples used by older logicians in dealing with this point are usually unfortunately chosen. They tend to pick out tables and chairs and apples on tables as test objects to demonstrate the object-like nature of reality and its one-to-one correspondence with logic. Man's artifacts and the agricultural products he severs from living plants have a unique degree of isolation; we may expect that languages will have fairly isolated terms for them. The real question is: What do different languages do, not with these artificially isolated objects but with the flowing face of nature in its motion, color, and changing form; with clouds, beaches, and yonder flight of birds? For, as goes our segmentation of the face of nature, so goes our physics of the Cosmos.

Here we find differences in segmentation and selection of basic terms. We might isolate something in nature by saying 'It is a dripping spring.' Apache

erects the statement on a verb *ga*: 'be white (including clear, uncolored, and so on).' With a prefix *nō-* the meaning of downward motion enters: 'whiteness moves downward.' Then *tó*, meaning both 'water' and 'spring' is prefixed. The result corresponds to our 'dripping spring,' but synthetically it is 'as water, or springs, whiteness moves downward.' How utterly unlike our way of thinking! The same verb, *ga*, with a prefix that means 'a place manifests the condition' becomes *gohlga*: 'the place is white, clear; a clearing, a plain.' These examples show that some languages have means of expression—chemical combination, as I called it—in which the separate terms are not so separate as in English but flow together into plastic synthetic creations. Hence such languages, which do not paint the separate-object picture of the universe to the same degree as English and its sister tongues, point toward possible new types of logic and possible new cosmical pictures.

The Indo-European languages and many others give great prominence to a type of sentence having two parts, each part built around a class of word—substantives and verbs—which those languages treat differently in grammar. As I showed in the April 1940 *Review*, this distinction is not drawn from nature; it is just a result of the fact that every tongue must have some kind of structure, and those tongues have made a go of exploiting this kind. The Greeks, especially Aristotle, built up this contrast and made it a law of reason. Since then, the contrast has been stated in logic in many different ways: subject and predicate, actor and action, things and relations between things, objects and their attributes, quantities and operations. And, pursuant again to grammar, the notion became ingrained that one of these classes of entities can exist in its own right but that the verb class cannot exist without an entity of the other class, the "thing" class, as a peg to hang on. "Embodiment is necessary," the watchword of this ideology, is seldom *strongly* questioned. Yet the whole trend of modern physics, with its emphasis on "the field," is an implicit questioning of the ideology. This contrast crops out in our mathematics as two kinds of symbols—the kind like $1, 2, 3, x, y, z$ and the kind like $+, -, \div, \sqrt{\ }$, log—though, in view of $0, \frac{1}{2}, \frac{3}{4}, \pi$, and others, perhaps no strict two-group classification holds. The two-group notion, however, is always present at the back of the thinking, although often not overtly expressed.

Our Indian languages show that with a suitable grammar we may have intelligent sentences that cannot be broken into subjects and predicates. Any attempted breakup is a breakup of some English translation or paraphrase of the sentence, not of the Indian sentence itself. We might as well try to decompose a certain synthetic resin into Celluloid and whiting because the resin can be imitated with Celluloid and whiting. The Algonkian language family, to which Shawnee belongs, does use a type of sentence like our subject and predicate but also gives prominence to the type shown by our examples in the text and in figure 31.1. To be sure, *ni-* is represented by a subject in the translation but means 'my' as well as 'I,' and the sentence could be translated thus: 'My hand is pulling the branch aside.' Or *ni-* might be absent; if so, we should be apt to manufacture a subject, like 'he, it, somebody,' or we could pick out for our English subject an idea corresponding to any one of the Shawnee elements.

When we come to Nootka, the sentence without subject or predicate is the only type. The term "predication" is used, but it means "sentence." Nootka has no parts of speech; the simplest utterance is a sentence, treating of some event

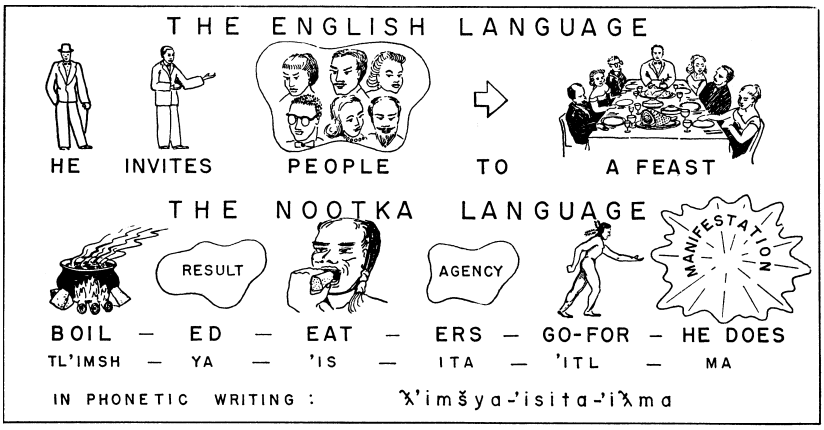


Figure 31.3
Here are shown the different ways in which English and Nootka formulate the same event. The English sentence is divisible into subject and predicate; the Nootka sentence is not, yet it is complete and logical. Furthermore, the Nootka sentence is just one word, consisting of the root *tl'imsh* with five suffixes.

or event-complex. Long sentences are sentences of sentences (complex sentences), not just sentences of words. In figure 31.3 we have a simple, not a complex, Nootka sentence. The translation, 'he invites people to a feast,' splits into subject and predicate. Not so the native sentence. It begins with the event of 'boiling or cooking,' *tl'imsh*; then comes *-ya* ('result') = 'cooked'; then *-'is* 'eating' = 'eating cooked food'; then *-ita* ('those who do') = 'eaters of cooked food'; then *-'itl* ('going for'); then *-ma*, sign of third-person indicative, giving *tl'imshya'isita'itlma*, which answers to the crude paraphrase, 'he, or somebody, goes for (invites) eaters of cooked food.'

The English technique of talking depends on the contrast of two artificial classes, substantives and verbs, and on the bipartitioned ideology of nature, already discussed. Our normal sentence, unless imperative, must have some substantive before its verb, a requirement that corresponds to the philosophical and also naïve notion of an actor who produces an action. This last might not have been so if English had had thousands of verbs like 'hold,' denoting positions. But most of our verbs follow a type of segmentation that isolates from nature what we call "actions," that is, moving outlines.

Following majority rule, we therefore read action into every sentence, even into 'I hold it.' A moment's reflection will show that 'hold' is no action but a state of relative positions. Yet we think of it and even see it as an action because language formulates it in the same way as it formulates more numerous expressions, like 'I strike it,' which deal with movements and changes.

We are constantly reading into nature fictional acting entities, simply because our verbs must have substantives in front of them. We have to say 'It flashed' or 'A light flashed,' setting up an actor, 'it' or 'light,' to perform what we call an action, "to flash." Yet the flashing and the light are one and the same! The Hopi language reports the flash with a simple verb, *rehpi*: 'flash (occurred).' There is no division into subject and predicate, not even a suffix like *-t* of Latin *tona-t* 'it

thunders.' Hopi can and does have verbs without subjects, a fact which may give that tongue potentialities, probably never to be developed, as a logical system for understanding some aspects of the universe. Undoubtedly modern science, strongly reflecting western Indo-European tongues, often does as we all do, sees actions and forces where it sometimes might be better to see states. On the other hand, 'state' is a noun, and as such it enjoys the superior prestige traditionally attaching to the subject or thing class; therefore science is exceedingly ready to speak of states if permitted to manipulate the concept like a noun. Perhaps, in place of the 'states' of an atom or a dividing cell, it would be better if we could manipulate as readily a more verblike concept but without the concealed premises of actor and action.

I can sympathize with those who say, "Put it into plain, simple English," especially when they protest against the empty formalism of loading discourse with pseudolearned words. But to restrict thinking to the patterns merely of English, and especially to those patterns which represent the acme of plainness in English, is to lose a power of thought which, once lost, can never be regained. It is the "plainest" English which contains the greatest number of unconscious assumptions about nature. This is the trouble with schemes like Basic English, in which an eviscerated British English, with its concealed premises working harder than ever, is to be fobbed off on an unsuspecting world as the substance of pure Reason itself. We handle even our plain English with much greater effect if we direct it from the vantage point of a multilingual awareness. For this reason I believe that those who envision a future world speaking only one tongue, whether English, German, Russian, or any other, hold a misguided ideal and would do the evolution of the human mind the greatest disservice. Western culture has made, through language, a provisional analysis of reality and, without correctives, holds resolutely to that analysis as final. The only correctives lie in all those other tongues which by aeons of independent evolution have arrived at different, but equally logical, provisional analyses.

In a valuable paper, "Modern logic and the task of the natural sciences," Harold N. Lee says: "Those sciences whose data are subject to quantitative measurement have been most successfully developed because we know so little about order systems other than those exemplified in mathematics. We can say with certainty, however, that there are other kinds, for the advance of logic in the last half century has clearly indicated it. We may look for advances in many lines in sciences at present well founded if the advance of logic furnishes adequate knowledge of other order types. We may also look for many subjects of inquiry whose methods are not strictly scientific at the present time to become so when new order systems are available."[1] To which may be added that an important field for the working out of new order systems, akin to, yet not identical with, present mathematics, lies in more penetrating investigation than has yet been made of languages remote in type from our own.

Notes

Reprinted from *Technol. Rev.*, 43:250–252, 266, 268, 272 (April 1941).

1. *Sigma Xi Quart.*, 28:125 (Autumn 1940).

PART XVI

Language 3—Pragmatics

Chapter 32

Logic and Conversation

H. P. Grice

It is a commonplace of philosophical logic that there are, or appear to be, divergences in meaning between, on the one hand, at least some of what I shall call the formal devices—$\sim$, $\wedge$, $\vee$, $\supset$, $(\forall x)$, $(\exists x)$, (ιx) (when these are given a standard two-valued interpretation)—and, on the other, what are taken to be their analogues or counterparts in natural language—such expressions as *not, and, or, if, all, some* (or *at least one*), *the*. Some logicians may at some time have wanted to claim that there are in fact no such divergences; but such claims, if made at all, have been somewhat rashly made, and those suspected of making them have been subjected to some pretty rough handling.

Those who concede that such divergences exist adhere, in the main, to one or the other of two rival groups, which I shall call the formalist and the informalist groups. An outline of a not uncharacteristic formalist position may be given as follows: Insofar as logicians are concerned with the formulation of very general patterns of valid inference, the formal devices possess a decisive advantage over their natural counterparts. For it will be possible to construct in terms of the formal devices a system of very general formulas, a considerable number of which can be regarded as, or are closely related to, patterns of inferences the expression of which involves some or all of the devices: Such a system may consist of a certain set of simple formulas that must be acceptable if the devices have the meaning that has been assigned to them, and an indefinite number of further formulas, many of which are less obviously acceptable and each of which can be shown to be acceptable if the members of the original set are acceptable. We have, thus, a way of handling dubiously acceptable patterns of inference, and if, as is sometimes possible, we can apply a decision procedure, we have an even better way. Furthermore, from a philosophical point of view, the possession by the natural counterparts of those elements in their meaning, which they do not share with the corresponding formal devices, is to be regarded as an imperfection of natural languages; the elements in question are undesirable excrescences. For the presence of these elements has the result both that the concepts within which they appear cannot be precisely or clearly defined, and that at least some statements involving them cannot, in some circumstances, be assigned a definite truth value; and the indefiniteness of these concepts not only is objectionable in itself but also leaves open the way to metaphysics—we cannot be certain that none of these natural language expressions is metaphysically "loaded." For these reasons, the expressions, as

From chapter 2 in *Syntax and Semantics 3: Speech Acts*, ed. P. Cole and J. Morgan (New York: Academic Press, 1975), 26–40. Reprinted with permission.

used in natural speech, cannot be regarded as finally acceptable, and may turn out to be, finally, not fully intelligible. The proper course is to conceive and begin to construct an ideal language, incorporating the formal devices, the sentences of which will be clear, determinate in truth value, and certifiably free from metaphysical implications; the foundations of science will now be philosophically secure, since the statements of the scientist will be expressible (though not necessarily actually expressed) within this ideal language. (I do not wish to suggest that all formalists would accept the whole of this outline, but I think that all would accept at least some part of it.)

To this, an informalist might reply in the following vein. The philosophical demand for an ideal language rests on certain assumptions that should not be conceded; these are, that the primary yardstick by which to judge the adequacy of a language is its ability to serve the needs of science, that an expression cannot be guaranteed as fully intelligible unless an explication or analysis of its meaning has been provided, and that every explication or analysis must take the form of a precise definition that is the expression or assertion of a logical equivalence. Language serves many important purposes besides those of scientific inquiry; we can know perfectly well what an expression means (and so a fortiori that it is intelligible) without knowing its analysis, and the provision of an analysis may (and usually does) consist in the specification, as generalized as possible, of the conditions that count for or against the applicability of the expression being analyzed. Moreover, while it is no doubt true that the formal devices are especially amenable to systematic treatment by the logician, it remains the case that there are very many inferences and arguments, expressed in natural language and not in terms of these devices, which are nevertheless recognizably valid. So there must be a place for an unsimplified, and so more or less unsystematic, logic of the natural counterparts of these devices; this logic may be aided and guided by the simplified logic of the formal devices but cannot be supplanted by it. Indeed, not only do the two logics differ, but sometimes they come into conflict; rules that hold for a formal device may not hold for its natural counterpart.

On the general question of the place in philosophy of the reformation of natural language, I shall, in this essay, have nothing to say. I shall confine myself to the dispute in its relation to the alleged divergences. I have, moreover, no intention of entering the fray on behalf of either contestant. I wish, rather, to maintain that the common assumption of the contestants that the divergences do in fact exist is (broadly speaking) a common mistake, and that the mistake arises from inadequate attention to the nature and importance of the conditions governing conversation. I shall, therefore, inquire into the general conditions that, in one way or another, apply to conversation as such, irrespective of its subject matter. I begin with a characterization of the notion of "implicature."

Implicature

Suppose that A and B are talking about a mutual friend, C, who is now working in a bank. A asks B how C is getting on in his job, and B replies, *Oh quite well, I think; he likes his colleagues, and he hasn't been to prison yet*. At this point, A

might well inquire what B was implying, what he was suggesting, or even what he meant by saying that C had not yet been to prison. The answer might be any one of such things as that C is the sort of person likely to yield to the temptation provided by his occupation, that C's colleagues are really very unpleasant and treacherous people, and so forth. It might, of course, be quite unnecessary for A to make such an inquiry of B, the answer to it being, in the context, clear in advance. It is clear that whatever B implied, suggested, meant in this example, is distinct from what B said, which was simply that C had not been to prison yet. I wish to introduce, as terms of art, the verb *implicate* and the related nouns *implicature* (cf. *implying*) and *implicatum* (cf. *what is implied*). The point of this maneuver is to avoid having, on each occasion, to choose between this or that member of the family of verbs for which *implicate* is to do general duty. I shall, for the time being at least, have to assume to a considerable extent an intuitive understanding of the meaning of *say* in such contexts, and an ability to recognize particular verbs as members of the family with which *implicate* is associated. I can, however, make one or two remarks that may help to clarify the more problematic of these assumptions, namely, that connected with the meaning of the word *say*.

In the sense in which I am using the word *say*, I intend what someone has said to be closely related to the conventional meaning of the words (the sentence) he has uttered. Suppose someone to have uttered the sentence *He is in the grip of a vice*. Given a knowledge of the English language, but no knowledge of the circumstances of the utterance, one would know something about what the speaker had said on the assumption that he was speaking standard English, and speaking literally. One would know that he had said, about some particular male person or animal *x*, that at the time of the utterance (whatever that was), either (1) *x* was unable to rid himself of a certain kind of bad character trait or (2) some part of *x*'s person was caught in a certain kind of tool or instrument (approximate account, of course). But for a full identification of what the speaker had said, one would need to know (a) the identity of *x*, (b) the time of utterance, and (c) the meaning, on the particular occasion of utterance, of the phrase *in the grip of a vice* [a decision between (1) and (2)]. This brief indication of my use of *say* leaves it open whether a man who says (today) *Harold Wilson is a great man* and another who says (also today) *The British Prime Minister is a great man* would, if each knew that the two singular terms had the same reference, have said the same thing. But whatever decision is made about this question, the apparatus that I am about to provide will be capable of accounting for any implicatures that might depend on the presence of one rather than another of these singular terms in the sentence uttered. Such implicatures would merely be related to different maxims.

In some cases the conventional meaning of the words used will determine what is implicated, besides helping to determine what is said. If I say (smugly), *He is an Englishman; he is, therefore, brave*, I have certainly committed myself, by virtue of the meaning of my words, to its being the case that his being brave is a consequence of (follows from) his being an Englishman. But while I have said that he is an Englishman, and said that he is brave, I do not want to say that I have *said* (in the favored sense) that it follows from his being an Englishman

that he is brave, though I have certainly indicated, and so implicated, that this is so. I do not want to say that my utterance of this sentence would be, *strictly speaking*, false should the consequence in question fail to hold. So *some* implicatures are conventional, unlike the one with which I introduced this discussion of implicature.

I wish to represent a certain subclass of nonconventional implicatures, which I shall call *conversational* implicatures, as being essentially connected with certain general features of discourse; so my next step is to try to say what these features are. The following may provide a first approximation to a general principle. Our talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts; and each participant recognizes in them, to some extent, a common purpose or set of purposes, or at least a mutually accepted direction. This purpose or direction may be fixed from the start (e.g., by an initial proposal of a question for discussion), or it may evolve during the exchange; it may be fairly definite, or it may be so indefinite as to leave very considerable latitude to the participants (as in a casual conversation). But at each stage, *some* possible conversational moves would be excluded as conversationally unsuitable. We might then formulate a rough general principle which participants will be expected (ceteris paribus) to observe, namely: Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. One might label this the Cooperative Principle.

On the assumption that some such general principle as this is acceptable, one may perhaps distinguish four categories under one or another of which will fall certain more specific maxims and submaxims, the following of which will, in general, yield results in accordance with the Cooperative Principle. Echoing Kant, I call these categories Quantity, Quality, Relation, and Manner. The category of Quantity relates to the quantity of information to be provided, and under it fall the following maxims:

1. Make your contribution as informative as is required (for the current purposes of the exchange).
2. Do not make your contribution more informative than is required.

(The second maxim is disputable; it might be said that to be overinformative is not a transgression of the Cooperative Principle but merely a waste of time. However, it might be answered that such overinformativeness may be confusing in that it is liable to raise side issues; and there may also be an indirect effect, in that the hearers may be misled as a result of thinking that there is some particular *point* in the provision of the excess of information. However this may be, there is perhaps a different reason for doubt about the admission of this second maxim, namely, that its effect will be secured by a later maxim, which concerns relevance.)

Under the category of Quality falls a supermaxim—"Try to make your contribution one that is true"—and two more specific maxims:

1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

Under the category of Relation I place a single maxim, namely, "Be relevant." Though the maxim itself is terse, its formulation conceals a number of problems that exercise me a good deal: questions about what different kinds and focuses of relevance there may be, how these shift in the course of a talk exchange, how to allow for the fact that subjects of conversation are legitimately changed, and so on. I find the treatment of such questions exceedingly difficult, and I hope to revert to them in later work.

Finally, under the category of Manner, which I understand as relating not (like the previous categories) to what is said but, rather, to *how* what is said is to be said, I include the supermaxim—"Be perspicuous"—and various maxims such as:

1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary prolixity).
4. Be orderly.

And one might need others.

It is obvious that the observance of some of these maxims is a matter of less urgency than is the observance of others; a man who has expressed himself with undue prolixity would, in general, be open to milder comment than would a man who has said something he believes to be false. Indeed, it might be felt that the importance of at least the first maxim of Quality is such that it should not be included in a scheme of the kind I am constructing; other maxims come into operation only on the assumption that this maxim of Quality is satisfied. While this may be correct, so far as the generation of implicatures is concerned it seems to play a role not totally different from the other maxims, and it will be convenient, for the present at least, to treat it as a member of the list of maxims.

There are, of course, all sorts of other maxims (aesthetic, social, or moral in character), such as "Be polite," that are also normally observed by participants in talk exchanges, and these may also generate nonconventional implicatures. The conversational maxims, however, and the conversational implicatures connected with them, are specially connected (I hope) with the particular purposes that talk (and so, talk exchange) is adapted to serve and is primarily employed to serve. I have stated my maxims as if this purpose were a maximally effective exchange of information; this specification is, of course, too narrow, and the scheme needs to be generalized to allow for such general purposes as influencing or directing the actions of others.

As one of my avowed aims is to see talking as a special case or variety of purposive, indeed rational, behavior, it may be worth noting that the specific expectations or presumptions connected with at least some of the foregoing maxims have their analogues in the sphere of transactions that are not talk exchanges. I list briefly one such analogue for each conversational category.

1. *Quantity.* If you are assisting me to mend a car, I expect your contribution to be neither more nor less than is required. If, for example, at a particular stage I need four screws, I expect you to hand me four, rather than two or six.

2. *Quality.* I expect your contributions to be genuine and not spurious. If I need sugar as an ingredient in the cake you are assisting me to make, I do not expect you to hand me salt; if I need a spoon, I do not expect a trick spoon made of rubber.
3. *Relation.* I expect a partner's contribution to be appropriate to the immediate needs at each stage of the transaction. If I am mixing ingredients for a cake, I do not expect to be handed a good book, or even an oven cloth (though this might be an appropriate contribution at a later stage).
4. *Manner.* I expect a partner to make it clear what contribution he is making and to execute his performance with reasonable dispatch.

These analogies are relevant to what I regard as a fundamental question about the Cooperative Principle and its attendant maxims, namely, what the basis is for the assumption which we seem to make, and on which (I hope) it will appear that a great range of implicatures depends, that talkers will in general (ceteris paribus and in the absence of indications to the contrary) proceed in the manner that these principles prescribe. A dull but, no doubt at a certain level, adequate answer is that it is just a well-recognized empirical fact that people do behave in these ways; they learned to do so in childhood and have not lost the habit of doing so; and, indeed, it would involve a good deal of effort to make a radical departure from the habit. It is much easier, for example, to tell the truth than to invent lies.

I am, however, enough of a rationalist to want to find a basis that underlies these facts, undeniable though they may be; I would like to be able to think of the standard type of conversational practice not merely as something that all or most do *in fact* follow but as something that it is *reasonable* for us to follow, that we *should not* abandon. For a time, I was attracted by the idea that observance of the Cooperative Principle and the maxims, in a talk exchange, could be thought of as a quasi-contractual matter, with parallels outside the realm of discourse. If you pass by when I am struggling with my stranded car, I no doubt have some degree of expectation that you will offer help, but once you join me in tinkering under the hood, my expectations become stronger and take more specific forms (in the absence of indications that you are merely an incompetent meddler); and talk exchanges seemed to me to exhibit, characteristically, certain features that jointly distinguish cooperative transactions:

1. The participants have some common immediate aim, like getting a car mended; their ultimate aims may, of course, be independent and even in conflict—each may want to get the car mended in order to drive off, leaving the other stranded. In characteristic talk exchanges, there is a common aim even if, as in an over-the-wall chat, it is a second-order one, namely, that each party should, for the time being, identify himself with the transitory conversational interests of the other.
2. The contributions of the participants should be dovetailed, mutually dependent.
3. There is some sort of understanding (which may be explicit but which is often tacit) that, other things being equal, the transaction should continue in appropriate style unless both parties are agreeable that it should terminate. You do not just shove off or start doing something else.

But while some such quasi-contractual basis as this may apply to some cases, there are too many types of exchange, like quarreling and letter writing, that it fails to fit comfortably. In any case, one feels that the talker who is irrelevant or obscure has primarily let down not his audience but himself. So I would like to be able to show that observance of the Cooperative Principle and maxims is reasonable (rational) along the following lines: that anyone who cares about the goals that are central to conversation/communication (such as giving and receiving information, influencing and being influenced by others) must be expected to have an interest, given suitable circumstances, in participation in talk exchanges that will be profitable only on the assumption that they are conducted in general accordance with the Cooperative Principle and the maxims. Whether any such conclusion can be reached, I am uncertain; in any case, I am fairly sure that I cannot reach it until I am a good deal clearer about the nature of relevance and of the circumstances in which it is required.

It is now time to show the connection between the Cooperative Principle and maxims, on the one hand, and conversational implicature on the other.

A participant in a talk exchange may fail to fulfill a maxim in various ways, which include the following:

1. He may quietly and unostentatiously *violate* a maxim; if so, in some cases he will be liable to mislead.
2. He may *opt out* from the operation both of the maxim and of the Cooperative Principle; he may say, indicate, or allow it to become plain that he is unwilling to cooperate in the way the maxim requires. He may say, for example, *I cannot say more; my lips are sealed*.
3. He may be faced by a *clash:* He may be unable, for example, to fulfill the first maxim of Quantity (Be as informative as is required) without violating the second maxim of Quality (Have adequate evidence for what you say).
4. He may *flout* a maxim; that is, he may blatantly fail to fulfill it. On the assumption that the speaker is able to fulfill the maxim and to do so without violating another maxim (because of a clash), is not opting out, and is not, in view of the blatancy of his performance, trying to mislead, the hearer is faced with a minor problem: How can his saying what he did say be reconciled with the supposition that he is observing the overall Cooperative Principle? This situation is one that characteristically gives rise to a conversational implicature; and when a conversational implicature is generated in this way, I shall say that a maxim is being *exploited*.

I am now in a position to characterize the notion of conversational implicature. A man who, by (in, when) saying (or making as if to say) that *p* has implicated that *q*, may be said to have conversationally implicated that *q*, provided that (1) he is to be presumed to be observing the conversational maxims, or at least the Cooperative Principle; (2) the supposition that he is aware that, or thinks that, *q* is required in order to make his saying or making as if to say *p* (or doing so in *those* terms) consistent with this presumption; and (3) the speaker thinks (and would expect the hearer to think that the speaker thinks) that it is within the competence of the hearer to work out, or grasp intuitively, that the supposition mentioned in (2) is required. Apply this to my initial

example, to B's remark that C has not yet been to prison. In a suitable setting A might reason as follows: "(1) B has apparently violated the maxim 'Be relevant' and so may be regarded as having flouted one of the maxims conjoining perspicuity, yet I have no reason to suppose that he is opting out from the operation of the Cooperative Principle; (2) given the circumstances, I can regard his irrelevance as only apparent if, and only if, I suppose him to think that C is potentially dishonest; (3) B knows that I am capable of working out step (2). So B implicates that C is potentially dishonest."

The presence of a conversational implicature must be capable of being worked out; for even if it can in fact be intuitively grasped, unless the intuition is replaceable by an argument, the implicature (if present at all) will not count as a conversational implicature; it will be a conventional implicature. To work out that a particular conversational implicature is present, the hearer will rely on the following data: (1) the conventional meaning of the words used, together with the identity of any references that may be involved; (2) the Cooperative Principle and its maxims; (3) the context, linguistic or otherwise, of the utterance; (4) other items of background knowledge; and (5) the fact (or supposed fact) that all relevant items falling under the previous headings are available to both participants and both participants know or assume this to be the case. A general pattern for the working out of a conversational implicature might be given as follows: "He has said that *p*; there is no reason to suppose that he is not observing the maxims, or at least the Cooperative Principle; he could not be doing this unless he thought that *q*; he knows (and knows that I know that he knows) that I can see that the supposition that he thinks that *q* is required; he has done nothing to stop me thinking that *q*; he intends me to think, or is at least willing to allow me to think, that *q*; and so he has implicated that *q*."

Examples of Conversational Implicature

I shall now offer a number of examples, which I shall divide into three groups.

Group A: *Examples in which no maxim is violated, or at least in which it is not clear that any maxim is violated*

A is standing by an obviously immobilized car and is approached by B; the following exchange takes place:

1. A: *I am out of petrol.*
   B: *There is a garage round the corner.*

(Gloss: B would be infringing the maxim "Be relevant" unless he thinks, or thinks it possible, that the garage is open, and has petrol to sell; so he implicates that the garage is, or at least may be open, etc.)

In this example, unlike the case of the remark *He hasn't been to prison yet*, the unstated connection between B's remark and A's remark is so obvious that, even if one interprets the supermaxim of Manner, "Be perspicuous," as applying not only to the expression of what is said but also to the connection of what is said with adjacent remarks, there seems to be no case for regarding that

supermaxim as infringed in this example. The next example is perhaps a little less clear in this respect:

2. A: *Smith doesn't seem to have a girlfriend these days.*
   B: *He has been paying a lot of visits to New York lately.*

B implicates that Smith has, or may have, a girlfriend in New York. (A gloss is unnecessary in view of that given for the previous example.)

In both examples, the speaker implicates that which he must be assumed to believe in order to preserve the assumption that he is observing the maxim of Relation.

Group B: *Examples in which a maxim is violated, but its violation is to be explained by the supposition of a clash with another maxim*

A is planning with B an itinerary for a holiday in France. Both know that A wants to see his friend C, if to do so would not involve too great a prolongation of his journey:

3. A: *Where does C live?*
   B: *Somewhere in the South of France.*

(Gloss: There is no reason to suppose that B is opting out; his answer is, as he well knows, less informative than is required to meet A's needs. This infringement of the first maxim of Quantity can be explained only by the supposition that B is aware that to be more informative would be to say something that infringed the second maxim of Quality. "Don't say what you lack adequate evidence for," so B implicates that he does not know in which town C lives.)

Group C: *Examples that involve exploitation, that is, a procedure by which a maxim is flouted for the purpose of getting in a conversational implicature by means of something of the nature of a figure of speech*

In these examples, though some maxim is violated at the level of what is said, the hearer is entitled to assume that that maxim, or at least the overall Cooperative Principle, is observed at the level of what is implicated.

(1a) *A flouting of the first maxim of Quantity*

A is writing a testimonial about a pupil who is a candidate for a philosophy job, and his letter reads as follows: "Dear Sir, Mr. X's command of English is excellent, and his attendance at tutorials has been regular. Yours, etc." (Gloss: A cannot be opting out, since if he wished to be uncooperative, why write at all? He cannot be unable, through ignorance, to say more, since the man is his pupil; moreover, he knows that more information than this is wanted. He must, therefore, be wishing to impart information that he is reluctant to write down. This supposition is tenable only if he thinks Mr. X is no good at philosophy. This, then, is what he is implicating.)

Extreme examples of a flouting of the first maxim of Quantity are provided by utterances of patent tautologies like *Women are women* and *War is war*. I would wish to maintain that at the level of what is said, in my favored sense, such remarks are totally noninformative and so, at that level, cannot but infringe the first maxim of Quantity in any conversational context. They are, of

course, informative at the level of what is implicated, and the hearer's identification of their informative content at this level is dependent on his ability to explain the speaker's selection of this particular patent tautology.

(1b) *An infringement of the second maxim of Quantity, "Do not give more information than is required," on the assumption that the existence of such a maxim should be admitted*

A wants to know whether *p*, and B volunteers not only the information that *p*, but information to the effect that it is certain that *p*, and that the evidence for its being the case that *p* is so-and-so and such-and-such.

B's volubility may be undesigned, and if it is so regarded by A it may raise in A's mind a doubt as to whether B is as certain as he says he is ("Methinks the lady doth protest too much"). But if it is thought of as designed, it would be an oblique way of conveying that it is to some degree controversial whether or not *p*. It is, however, arguable that such an implicature could be explained by reference to the maxim of Relation without invoking an alleged second maxim of Quantity.

(2a) *Examples in which the first maxim of Quality is flouted*

Irony. X, with whom A has been on close terms until now, has betrayed a secret of A's to a business rival. A and his audience both know this. A says *X is a fine friend*. (Gloss: It is perfectly obvious to A and his audience that what A has said or has made as if to say is something he does not believe, and the audience knows that A knows that this is obvious to the audience. So, unless A's utterance is entirely pointless, A must be trying to get across some other proposition than the one he purports to be putting forward. This must be some obviously related proposition; the most obviously related proposition is the contradictory of the one he purports to be putting forward.)

Metaphor. Examples like *You are the cream in my coffee* characteristically involve categorial falsity, so the contradictory of what the speaker has made as if to say will, strictly speaking, be a truism; so it cannot be *that* that such a speaker is trying to get across. The most likely supposition is that the speaker is attributing to his audience some feature or features in respect of which the audience resembles (more or less fancifully) the mentioned substance.

It is possible to combine metaphor and irony by imposing on the hearer two stages of interpretation. I say *You are the cream in my coffee*, intending the hearer to reach first the metaphor interpretant "You are my pride and joy" and then the irony interpretant "You are my bane."

Meiosis. Of a man known to have broken up all the furniture, one says *He was a little intoxicated*.

Hyperbole. Every nice girl loves a sailor.

(2b) Examples in which the second maxim of Quality, "Do not say that for which you lack adequate evidence," is flouted are perhaps not easy to find, but the following seems to be a specimen. I say of X's wife, *She is probably deceiving him this evening*. In a suitable context, or with a suitable gesture or tone of voice, it may be clear that I have no adequate reason for supposing this to be the case. My partner, to preserve the assumption that the conversational game is still being played, assumes that I am getting at some related proposition for the acceptance of which I do have a reasonable basis. The related proposition might

well be that she is given to deceiving her husband, or possibly that she is the sort of person who would not stop short of such conduct.

(3) *Examples in which an implicature is achieved by real, as distinct from apparent, violation of the maxim of Relation* are perhaps rare but the following seems to be a good candidate. At a genteel tea party, A says *Mrs. X is an old bag*. There is a moment of appalled silence, and then B says *The weather has been quite delightful this summer, hasn't it?* B has blatantly refused to make what he says relevant to A's preceding remark. He thereby implicates that A's remark should not be discussed and, perhaps more specifically, that A has committed a social gaffe.

(4) *Examples in which various maxims falling under the supermaxim "Be perspicuous" are flouted*

Ambiguity. We must remember that we are concerned only with ambiguity that is deliberate, and that the speaker intends or expects to be recognized by his hearer. The problem the hearer has to solve is why a speaker should, when still playing the conversational game, go out of his way to choose an ambiguous utterance. There are two types of cases:

(a) Examples in which there is no difference, or no striking difference, between two interpretations of an utterance with respect to straightforwardness; neither interpretation is notably more sophisticated, less standard, more recondite or more far-fetched than the other. We might consider Blake's lines: "Never seek to tell thy love, Love that never told can be." To avoid the complications introduced by the presence of the imperative mood, I shall consider the related sentence, *I sought to tell my love, love that never told can be*. There may be a double ambiguity here. *My love* may refer to either a state of emotion or an object of emotion, and *love that never told can be* may mean either "Love that cannot be told" or "love that if told cannot continue to exist." Partly because of the sophistication of the poet and partly because of internal evidence (that the ambiguity is kept up), there seems to be no alternative to supposing that the ambiguities are deliberate and that the poet is conveying both what he would be saying if one interpretation were intended rather than the other, and vice versa; though no doubt the poet is not explicitly saying any one of these things but only conveying or suggesting them (cf. "Since she [nature] pricked thee out for women's pleasure, mine be thy love, and thy love's use their treasure").

(b) Examples in which one interpretation is notably less straightforward than another. Take the complex example of the British General who captured the province of Sind and sent back the message *Peccavi*. The ambiguity involved ("I have Sind"/"I have sinned") is phonemic, not morphemic; and the expression actually used is unambiguous, but since it is in a language foreign to speaker and hearer, translation is called for, and the ambiguity resides in the standard translation into native English.

Whether or not the straightforward interpretant ("I have sinned") is being conveyed, it seems that the nonstraightforward interpretant must be. There might be stylistic reasons for conveying by a sentence merely its nonstraightforward interpretant, but it would be pointless, and perhaps also stylistically objectionable, to go to the trouble of finding an expression that nonstraightforwardly conveys that *p*, thus imposing on an audience the effort involved in finding this interpretant, if this interpretant were otiose so far as communication

was concerned. Whether the straightforward interpretant is also being conveyed seems to depend on whether such a supposition would conflict with other conversational requirements, for example, would it be relevant, would it be something the speaker could be supposed to accept, and so on. If such requirements are not satisfied, then the straightforward interpretant is not being conveyed. If they are, it is. If the author of *Peccavi* could naturally be supposed to think that he had committed some kind of transgression, for example, had disobeyed his orders in capturing Sind, and if reference to such a transgression would be relevant to the presumed interests of the audience, then he would have been conveying both interpretants: otherwise he would be conveying only the nonstraightforward one.

Obscurity. How do I exploit, for the purposes of communication, a deliberate and overt violation of the requirement that I should avoid obscurity? Obviously, if the Cooperative Principle is to operate, I must intend my partner to understand what I am saying despite the obscurity I import into my utterance. Suppose that A and B are having a conversation in the presence of a third party, for example, a child, then A might be deliberately obscure, though not too obscure, in the hope that B would understand and the third party not. Furthermore, if A expects B to see that A is being deliberately obscure, it seems reasonable to suppose that, in making his conversational contribution in this way, A is implicating that the contents of his communication should not be imparted to the third party.

Failure to be brief or succinct. Compare the remarks:

(a) *Miss X sang "Home Sweet Home."*
(b) *Miss X produced a series of sounds that corresponded closely with the score of "Home Sweet Home."*

Suppose that a reviewer has chosen to utter (b) rather than (a). (Gloss: Why has he selected that rigmarole in place of the concise and nearly synonymous *sang?* Presumably, to indicate some striking difference between Miss X's performance and those to which the word *singing* is usually applied. The most obvious supposition is that Miss X's performance suffered from some hideous defect. The reviewer knows that this supposition is what is likely to spring to mind, so that is what he is implicating.)

Generalized Conversational Implicature

I have so far considered only cases of what I might call "particularized conversational implicature"—that is to say, cases in which an implicature is carried by saying that *p* on a particular occasion in virtue of special features of the context, cases in which there is no room for the the idea that an implicature of this sort is normally carried by saying that *p*. But there are cases of generalized conversational implicature. Sometimes one can say that the use of a certain form of words in an utterance would normally (in the absence of special circumstances) carry such-and-such an implicature or type of implicature. Noncontroversial examples are perhaps hard to find, since it is all too easy to treat a generalized conversational implicature as if it were a conventional implicature. I offer an example that I hope may be fairly noncontroversial.

Anyone who uses a sentence of the form *X is meeting a woman this evening* would normally implicate that the person to be met was someone other than X's wife, mother, sister, or perhaps even close platonic friend. Similarly, if I were to say *X went into a house yesterday and found a tortoise inside the front door,* my hearer would normally be surprised if some time later I revealed that the house was X's own. I could produce similar linguistic phenomena involving the expressions *a garden, a car, a college*, and so on. Sometimes, however, there would normally be no such implicature ("I have been sitting in a car all morning"), and sometimes a reverse implicature ("I broke a finger yesterday"). I am inclined to think that one would not lend a sympathetic ear to a philosopher who suggested that there are three senses of the form of expression *an X:* one in which it means roughly "something that satisfies the conditions defining the word X," another in which it means approximately "an X (in the first sense) that is only remotely related in a certain way to some person indicated by the context," and yet another in which it means "an X (in the first sense) that is closely related in a certain way to some person indicated by the context." Would we not much prefer an account on the following lines (which, of course, may be incorrect in detail): When someone, by using the form of expression *an X*, implicates that the X does not belong to or is not otherwise closely connected with some identifiable person, the implicature is present because the speaker has failed to be specific in a way in which he might have been expected to be specific, with the consequence that it is likely to be assumed that he is not in a position to be specific. This is a familiar implicature situation and is classifiable as a failure, for one reason or another, to fulfill the first maxim of Quantity. The only difficult question is why it should, in certain cases, be presumed, independently of information about particular contexts of utterance, that specification of the closeness or remoteness of the connection between a particular person or object and a further person who is mentioned or indicated by the utterance should be likely to be of interest. The answer must lie in the following region: Transactions between a person and other persons or things closely connected with him are liable to be very different as regards their concomitants and results from the same sort of transactions involving only remotely connected persons or things; the concomitants and results, for instance, of my finding a hole in my roof are likely to be very different from the concomitants and results of my finding a hole in someone else's roof. Information, like money, is often given without the giver's knowing to just what use the recipient will want to put it. If someone to whom a transaction is mentioned gives it further consideration, he is likely to find himself wanting the answers to further questions that the speaker may not be able to identify in advance; if the appropriate specification will be likely to enable the hearer to answer a considerable variety of such questions for himself, then there is a presumption that the speaker should include it in his remark; if not, then there is no such presumption.

Finally, we can now show that, conversational implicature being what it is, it must possess certain features:

1. Since, to assume the presence of a conversational implicature, we have to assume that at least the Cooperative Principle is being observed, and since it is possible to opt out of the observation of this principle, it follows that a

generalized conversational implicature can be canceled in a particular case. It may be explicitly canceled, by the addition of a clause that states or implies that the speaker has opted out, or it may be contextually canceled, if the form of utterance that usually carries it is used in a context that makes it clear that the speaker is opting out.

2. Insofar as the calculation that a particular conversational implicature is present requires, besides contextual and background information, only a knowledge of what has been said (or of the conventional commitment of the utterance), and insofar as the manner of expression plays no role in the calculation, it will not be possible to find another way of saying the same thing, which simply lacks the implicature in question, except where some special feature of the substituted version is itself relevant to the determination of an implicature (in virtue of one of the maxims of Manner). If we call this feature nondetachability, one may expect a generalized conversational implicature that is carried by a familiar, nonspecial locution to have a high degree of nondetachability.

3. To speak approximately, since the calculation of the presence of a conversational implicature presupposes an initial knowledge of the conventional force of the expression the utterance of which carries the implicature, a conversational implicatum will be a condition that is not included in the original specification of the expression's conventional force. Though it may not be impossible for what starts life, so to speak, as a conversational implicature to become conventionalized, to suppose that this is so in a given case would require special justification. So, initially at least, conversational implicata are not part of the meaning of the expressions to the employment of which they attach.

4. Since the truth of a conversational implicatum is not required by the truth of what is said (what is said may be true—what is implicated may be false), the implicature is not carried by what is said, but only by the saying of what is said, or by "putting it that way."

5. Since, to calculate a conversational implicature is to calculate what has to be supposed in order to preserve the supposition that the Cooperative Principle is being observed, and since there may be various possible specific explanations, a list of which may be open, the conversational implicatum in such cases will be disjunction of such specific explanations; and if the list of these is open, the implicatum will have just the kind of indeterminacy that many actual implicata do in fact seem to possess.

Chapter 33

Idiomaticity and Human Cognition

Raymond W. Gibbs Jr.

Figurative language has finally become a respectable area of study in the cognitive sciences. Most of the emphasis in this research effort has been on the interpretation of metaphor. However, idiomaticity has recently become a significant topic of concern in psycholinguistics, linguistics, developmental psychology, neuropsychology, and computer science (cf. Cacciari & Tabossi 1993). This interest in idiomaticity is well founded, given that American English, for example, contains many thousands of formulaic phrases and expressions that the ordinary speaker must somehow learn (as is evident in the many idiom and slang dictionaries currently available).

People are not considered competent speakers of a language until they master the various clichéd, idiomatic expressions that are ubiquitous in everyday discourse. Consider for a moment the following idiomatic expressions that are currently used by American college students (Munro 1989).

(1) a. From the way he was eyeing that girl, it was obvious that he was going to bust a move.
b. My friends rampaged through the kitchen when they got the munchies.
c. I thought he was so handsome that I wanted to jump his bones.
d. My boss was really upset and I wish he'd take a chill pill.

Do you understand what phrases such as *bust a move, have the munchies*, or *take a chill pill* mean? Why do speakers create and use these particular phrases or even more common phrases such as *blow your stack, spill the beans, get pissed off, kick the bucket*, or *pop the question?* Most scholars traditionally assume that idioms like these may have once been metaphorical in their origins but have lost their metaphoricity over time and now exist in the speakers' mental lexicons as stock formulas or as "dead" metaphors. Just as speakers no longer view *face of the clock* or *arm of a chair* as metaphoric, few contemporary people recognize phrases such as *have the munchies* or *to get pissed off* as particularly creative or metaphoric. For this reason, idioms are mostly thought to have relatively simple interpretations and, unlike metaphors, do not resist paraphrase. We may not know exactly why idioms mean what they do, but we understand that idioms have brief, clear definitions.

At the same time, idiomatic phrases are traditionally seen as being distinct from ordinary literal language because they are noncompositional in that their

From chapter 5 in *Idioms: Structural and Psychological Perspectives*, ed. Martin Everaert et al. (Hillsdale, NJ: Lawrence Erlbaum Associates, 1995), 97–116. Reprinted with permission.

conventional interpretations are not functions of the meanings of their individual parts (Chafe 1970; Chomsky 1965, 1980; Fraser 1970; Katz 1973; Weinreich 1969). For instance, the conventional, nonliteral interpretations of *blow your stack* or *to get pissed off* (i.e., 'to get very angry') cannot be determined through an analysis of their individual word meanings. Many linguists have also noted that the noncompositional nature of idioms explains why idioms tend to be limited in their syntactic and lexical productivity. For example, one cannot syntactically transform the phrase *John kicked the bucket* into a passive construction (i.e., **The bucket was kicked by John*) without disrupting its nonliteral meaning. Similarly, the noncompositionality of idioms also explains why idioms are lexically frozen (i.e., why one cannot change *kick the bucket* into *kick the pail* without disrupting its figurative meaning of 'to die'). Finally, speakers learn the meanings of idioms by forming arbitrary links between idioms and their nonliteral meanings (e.g., forming links between *spill the beans* and 'to reveal a secret,' *button your lips* and 'keep a secret,' *lose your marbles* and 'go crazy'). Thus, children and second language learners presumably learn idioms in a rote manner or simply infer the meanings of idioms from context.

How do we comprehend what idioms mean? The noncompositional view of idioms suggests that idioms are understood through the retrieval of their stipulated meanings from the lexicon once their literal meanings have been rejected as inappropriate, or in parallel to the processing of their literal meanings, or directly without any analysis of their overall literal meanings as phrases. A variety of experimental studies in psycholinguistics indicate that figurative uses of idioms are easier to process than literal uses (cf. Gibbs 1980, 1985, 1986). The common explanation of this finding is that people do not perform normal analyses on the individual lexical items when understanding idiom phrases, an assumption that makes sense given the traditional view of idioms as having noncompositional meaning.

My aim in this chapter is to challenge many of these traditional assumptions about idiomaticity. I argue that most linguists and psychologists are simply wrong to assert that idioms are noncompositional and have meanings that are derived from dead metaphors. A great deal of evidence in linguistics and psychology shows that many idioms are, at least to some extent, compositional or analyzable. People do not simply assume that the meanings of idioms are arbitrary or fixed by convention. Instead, people make sense of idiomatic expressions precisely because of their ordinary metaphorical and, to a lesser extent, metonymic knowledge that provides part of the link between these phrases and their figurative interpretations. Research in cognitive linguistics and experimental psychology supports the idea that idioms retain much of their metaphoricity. Many idioms are partly motivated by pervasive, preexisting metaphorical concepts that can account for significant aspects of the linguistic behavior of idioms as well as for the acquisition and comprehension of idioms. The metaphorical mappings underlying many idiomatic phrases give rise to multiple entailments, one reason why people understand idioms as having complex interpretations, contrary to the traditional view of idiomaticity.

My most general claim in this chapter is that the empirical study of idiomaticity reveals important aspects of how people think and reason about the concepts to which idioms refer. In this way, the study of idiomaticity in language

provides significant insights into the fundamental figurative character of human cognition.

The Analyzability of Idioms

There are many problems with the traditional view of idioms. First, consider whether idioms are noncompositional. The traditional belief is that any expression whose meaning is not predictable from an analysis of the meanings of its parts must have an arbitrary, or unmotivated, interpretation. For instance, the classic example of *kick the bucket* (meaning 'die') must receive its meaning by arbitrary stipulation because the words *kick, the*, and *bucket* have little to do with the act of dying. Of course, there may be some obscure historical reason why people use *kick the bucket* to talk of dying, but contemporary speakers are often unsure, or even completely ignorant, of why this phrase means what it does.

One major difficulty with this analysis is that scholars tend to draw false generalizations from an analysis of a single example (e.g., *kick the bucket*) or from just a few idiomatic phrases. Even though *kick the bucket* nicely illustrates some of the traditional claims about idioms, it is not particularly representative of the many kinds of idioms in American English. As noted by an increasing number of idiom scholars, it is clearly problematic to assume that idioms form a homogeneous class of linguistic items. Careful attention must be paid to the many syntactic, lexical, semantic, and pragmatic differences that exist among words and phrases that are generally judged to be idiomatic (i.e., those listed in standard idiom dictionaries).

The investigation of a wide range of idioms clearly demonstrates that many idioms are analyzable and have figurative meanings that are at least partly motivated (Cacciari 1993; Cacciari & Glucksberg 1990; Fillmore, Kay, & O'Connor 1988; Gibbs 1992, 1993; Gibbs & Nayak 1989; Glucksberg 1993; Lakoff 1987; Langacker 1986; Nunberg 1978; Ruwet 1992; Wasow, Sag, & Nunberg 1983). That is, many idioms, perhaps thousands, have individual components that independently contribute to what these phrases figuratively mean as wholes. For example, speakers know that *spill the beans* is analyzable because *beans* refers to an idea or secret and *spilling* refers to the act of revealing the secret. Similarly, in the phrase *pop the question*, it is easy to discern that the noun *question* refers to a marriage proposal when the verb *pop* is used to refer to the act of uttering it. People recognize that *blow your stack* is analyzable because *blow* refers to the act of suddenly releasing or expressing internal pressure from the *stack* or from the human mind/body.

Idioms differ in the extent to which they are analyzable. Some expressions are almost completely compositional or analyzable (e.g., *pop the question* and *blow your stack*), whereas other phrases are much less analyzable (e.g., *chew the fat* and *kick the bucket*). Many idioms that seem analyzable are also capable of being syntactically and lexically altered (e.g., *leave no legal stone unturned, Your remark touched a nerve I didn't know existed,* or *He pulled a string or two to help you get the job*). Many idioms share similar linguistic properties, as do most literal expressions, whereas other phrases are more classically formulaic (but far fewer than most scholars imagine!). We may not be able to predict exactly what

an idiom means through an analysis of the meanings of its individual words, but we can do more than throw our hands up and simply assert that the meanings of idioms are arbitrary and noncompositional.

It is very important to understand that the independent meanings that the components in idioms contribute to their phrases' overall meanings are not necessarily their putative literal meanings. Thus, to say that phrases such as *blow your stack* or *spill the beans* are analyzable to some degree means only that their individual components contribute some sort of independent meanings to the phrases' overall interpretations. Listeners do not necessarily understand *blow* and *spill* or *stack* and *beans* in their most literal senses. Instead, we understand *spill* as independently referring to the act of revealing the *beans* or the set of ideas that are being held secret.

Most linguists and psychologists view the problem of idiom comprehension as one where a reader or listener encounters an idiom and at some point switches from a normal, literal mode of processing to a more specialized, nonliteral mode of processing (i.e., where the stipulated meaning of the phrase is directly retrieved from the lexicon). I reject this widely held belief. It is unclear whether people actually switch from one mode of processing to another during idiom comprehension. In the first place, there are good reasons to believe that it is nearly impossible even to state what a word or phrase literally means (Gibbs 1984, 1989, 1994). Even when parsing very literal expressions such as *The cat is on the mat*, it is by no means clear what constitutes the literal meanings of the words in this sentence or the literal meaning of the sentence as a whole (see Searle 1979).

As is the case with many figurative language researchers, most idiom scholars mistakenly assume that the literal meaning of any word or phrase can be uniquely determined and that the literal meanings of idioms can somehow be easily distinguished from their figurative or nonliteral meanings. For example, the contrast between idioms and their literal meanings, metaphors and their literal meanings, metonymies and their literal meanings, ironic statements and their literal meanings provides very different notions of literal meaning. Empirical studies show that people have multiple, often contradictory, concepts of literal meaning that are implicit in scholarly discussions of linguistic meaning and interpretation (e.g., literal meaning as conventional meaning, context-free meaning, truth-conditional meaning, subject-matter meaning) (Gibbs, Buchalter, Moise, & Farrar 1993; Lakoff 1986). People appear to apply different senses of the concept of literal meaning in different ways depending on the kind of utterance, the context, and the task. Moreover, many psychological studies demonstrate that people do not access the same invariant, literal meanings each and every time they encounter a word in spoken or written discourse (Gibbs 1994). In general, we cannot assume that the words in idioms or entire idiom phrases have easily determined literal meanings. We do not have an adequate sense of what is meant by literal meaning or word meaning (and these are different notions) to continue assuming that idiom processing begins with some literal analysis and then switches to a specialized idiom mode of understanding.

This conclusion about literality and idiomaticity may be disturbing to many language scholars who continue to adhere to the idea that word meanings can

be defined in some context-free manner. Idiom researchers face the challenge of understanding the exact contribution of lexical semantics in people's interpretation of idiomatic phrases. My suspicion is that idioms differ quite a bit in terms of how their parts contribute to their meaning as a whole (see Nunberg 1978). Thus, even two analyzable expressions, such as *pop the question* and *spill the beans*, differ in that the word *question* has a fairly conventional meaning in the context of the idiom (i.e., having to do with the question about marriage), whereas the word *beans* has little to do with its conventional meaning as a food item. Various empirical studies have begun to examine how the meanings of individual words, whether these are characterized as literal, metaphorical, conventional, or whatever, are accessed during the processing of familiar and unfamiliar idioms (Cacciari 1993; Flores d'Arcais 1993). Some of these studies have revealed that many idioms have key points or uniqueness points, places at which idioms become uniquely identifiable (Cacciari & Tabossi 1988; Tabossi & Zardon 1993). These different studies provide evidence that is consistent with the data on the analyzability of idioms. But much work needs to be done in both linguistics and psychology on the different types of word meanings that play a role in the linguistic behavior (e.g., syntactic productivity) and in the learning, use, and understanding of idioms.

There are a number of interesting linguistic and behavioral consequences of the idea that idioms differ in their degree of analyzability. One series of studies showed that the semantic analyzability of an idiom affects people's intuitions about its syntactic productivity (Gibbs & Nayak 1989). For instance, people find semantically analyzable or decomposable idioms more syntactically flexible than unanalyzable idioms. Thus, an analyzable phrase such as *John laid down the law* can be syntactically altered into *The law was laid down by John* without disrupting its figurative meaning. However, semantically unanalyzable idioms tend to be much more syntactically frozen (e.g., one cannot change *John kicked the bucket* into *The bucket was kicked by John* without disrupting its figurative meaning).

Another series of studies indicated that semantic analyzability influences people's intuitions about the lexical flexibility of idioms (Gibbs, Nayak, Bolton, & Keppel 1989). Thus, analyzable idioms can be lexically altered without significant disruption of their nonliteral meanings (e.g., *button your lips* to *fasten your lips*), but semantically unanalyzable phrases cannot (e.g., *kick the bucket* to *punt the bucket*). More dramatically, the individual words in many idioms can be changed to create new idiomatic meanings that are based on both the original idiom's meaning and the new words. For example, the idiom *break the ice* can be altered to form *shatter the ice*, which now has the meaning of something like 'break down an uncomfortable and stiff social situation flamboyantly in one fell swoop!' (McGlone, Glucksberg, & Cacciari 1994). McGlone et al. argued that *shatter the ice* is an example not of lexical flexibility but of semantic productivity. People can understand semantically productive idiom variants (e.g., *Sam didn't spill a single bean*) quite readily; and the more familiar the original idiom, the more comprehensible the variant. Variant idioms can also be understood as quickly as their literal paraphrases (e.g., *Sam didn't spill a single bean* versus *Sam didn't say a single word*) (McGlone, Glucksberg, & Cacciari 1994).

The results of these different psycholinguistic studies demonstrate that the syntactic versatility, lexical flexibility, and semantic productivity of idioms are not arbitrary phenomena, perhaps due to historical reasons, but can be at least partially explained in terms of an idiom's semantic analyzability.

The analyzability of idioms also plays an important role in their immediate online interpretations. Because the individual components in analyzable idioms (e.g., *lay down the law*) systematically contribute to the figurative meanings of these phrases, people process idioms in a compositional manner in which the meanings of the components are accessed and combined according to the syntactic rules of the language (Peterson & Burgess 1993). On the other hand, a strict compositional analysis of semantically unanalyzable idioms (e.g., *kick the bucket*) provides little information about the figurative meanings of these expressions. Understanding unanalyzable idioms requires that people first do some sort of analysis where the individual words are examined to see if they have independent meanings that contribute to the meaningful interpretation of the idiom as a whole. Once this process fails to produce an acceptable interpretation in a context, then people probably retrieve the conventional, figurative meanings of these phrases from their mental lexicons.

Support for this idea about idiom comprehension comes from reading-time studies that showed that people took significantly less time to process decomposable or analyzable idioms than to read unanalyzable expressions (Gibbs, Nayak, & Cutting 1989). These data suggest that people normally attempt to do some compositional analysis when understanding all types of idiomatic phrases. This does not mean, however, that people automatically compute the literal, context-free interpretations of idioms (Gibbs 1980, 1985, 1986). The results of one study, for instance, showed that literally ill-formed idioms (e.g., *pop the question*) are understood just as quickly as are well-formed phrases (e.g., *kick the question*) (Gibbs, Nayak, & Cutting 1989). Thus, people do not appear to be biased toward processing the putative literal meanings of idioms. Rather, some compositional process attempts to assign some context-sensitive meanings to the individual components in idioms during understanding. Children actually experience greater difficulty learning the meanings of semantically unanalyzable idioms precisely because these phrases' nonliteral interpretations cannot be determined through analyses of their individual parts (Gibbs 1987, 1991; Nippold & Martin 1989). These data show that idiom learning does not occur in a rote manner but develops in stages as children acquire linguistic and metalinguistic skills (Cacciari & Levorato 1989; Levorato 1993; Levorato & Cacciari 1992).

Once again, the traditional view that idioms are dead metaphors with noncompositional meaning cannot account for any of these empirical findings. Many linguistic studies indicate that analyzability is an important concept in understanding the linguistic behavior of many other formulaic expressions, including verb-particle constructions (Bolinger 1971; Lindner 1981) and binomial expressions (Lambrecht 1984). One future challenge will be to see whether these observations and empirical findings on idiom analyzability extend to languages other than English and to linguistic constructions that are not normally classified as idiomatic (cf. Coulmas 1981).

The Cognitive Motivation for Idiomatic Meaning

One interesting characteristic of idiomaticity is that most languages have many idioms with similar figurative meanings. For example, American English has many idioms referring to the concept of getting angry (e.g., *blow your stack, hit the ceiling, blow off steam, bite your head off, get pissed off*). Another example is that American speakers may use *spill the beans, let the cat out of the bag, blow the lid off*, or *blow the whistle* to convey the idea of revealing or exposing a secret. According to the traditional view of idioms, there is no particular reason why we might create and use so many different expressions to convey roughly the same idea or concept. Each phrase's meaning is supposedly determined by separate historical situations that have evolved into pragmatic conventions of use. Again, the link between an idiom and its figurative meaning is arbitrary and cannot be predicted from the meanings of its individual words.

Do people understand that idiomatic meanings are arbitrarily determined? Or is there some underlying motivation for the figurative meanings associated with idioms? Contrary to the traditional view, the figurative meanings of idioms might well be motivated by people's conceptual knowledge that is itself constituted by metaphor. For example, the idiom *John spilled the beans* maps our knowledge of someone tipping over a container of beans to that of a person revealing some previously hidden secret. English speakers understand *spill the beans* to mean 'reveal the secret' because there are underlying conceptual metaphors, such as THE MIND IS A CONTAINER and IDEAS ARE PHYSICAL ENTITIES, that structure their conceptions of minds, secrets, and disclosure (Lakoff & Johnson 1980). Even though the existence of these conceptual metaphors does not predict that certain idioms or conventional expressions must appear in the language (e.g., that we have the expression *spill the beans* as opposed to *spill the peas*), the presence of these independent conceptual metaphors by which we make sense of experience partially explains why specific phrases (e.g., *spill the beans*) are used to refer to particular events (e.g., the revealing of secrets).

My claim that idioms are partially motivated by conceptual metaphor contrasts with the traditional notion that idioms arise from dead metaphors. Scholars adhering to the traditional view confuse conventional with dead metaphors. They insist that idiomatic meaning arises mostly from historical circumstances that are opaque to contemporary speakers and have little to do with ordinary human cognition. But determining whether an idiom is dead or just conventional requires, among other things, a search for its systematic manifestation in the language as a whole and in our everyday reasoning patterns. One of the advantages of not simply looking at isolated examples but instead examining groups of idioms, especially those referring to similar concepts, is that it is easier to uncover the active presence of conceptual metaphors (i.e., metaphors that actively structure the way we think about different domains of experience). There are plenty of basic conventional metaphors that are alive, certainly enough to show that what is conventional and fixed need not be dead (Lakoff & Turner 1989). Part of the problem with the traditional view of idioms stems from its inability to reflect contemporary speakers' metaphorical

schemes of thought. For this reason, the traditional view simply cannot explain why the figurative meanings of so many idioms make sense to speakers.

Various researchers in cognitive linguistics have explored a large number of representative domains of human experience (e.g., time, causation, spatial orientation, ideas, anger, understanding) to demonstrate the pervasiveness of various metaphorical systems in our everyday thought, at least as these ideas are manifested in the language people use (Johnson 1987; Kovecses 1986; Lakoff 1987, 1990; Lakoff & Johnson 1980; Lakoff & Turner 1989; Sweetser 1990; Turner 1991). This work adheres to the commitment in cognitive linguistics that theories of linguistic structure and use must be in accord with what is generally known about the human mind from different disciplines in the cognitive sciences (Gibbs 1996; Lakoff 1990). This commitment entails the belief that the analysis of the conceptual and experiential basis of linguistic categories and constructs is of primary importance. For this reason, the formal structures of language are studied not as if they were autonomous, but as reflections of general conceptual organization, categorization principles, and processing mechanisms. By explicitly looking for links between linguistic structure and ordinary cognition, cognitive linguists do not take the risk, as do most linguists of the generative persuasion, of ignoring most influences of thought on language. This research strategy has been quite beneficial to our understanding of idioms as partly motivated, and not arbitrary, linguistic phenomena.

Some of the cognitive linguistic analyses of idioms provide some evidence for the idea that idioms do not exist as separate units within the lexicon but actually reflect coherent systems of metaphorical concepts (Kovecses 1986; Lakoff 1987). For example, the idiomatic phrases *blow your stack, flip your lid, hit the ceiling, get hot under the collar, lose your cool,* and *get steamed up* appear to be motivated by the conceptual metaphor ANGER IS HEATED FLUID IN A CONTAINER, which is one of the small set of conceptual mappings between different source and target domains that form part of our conceptualization for anger. These same conceptual mappings give rise to many of the conventional expressions that are often viewed as nonidiomatic (e.g., *I exploded with anger*).

But is there any evidence that conceptual metaphors, such as ANGER IS HEATED FLUID IN A CONTAINER, are really conceptual and not, more simply, generalizations of linguistic meaning? We might understand, for instance, that *blow your stack, flip your lid, hit the ceiling*, and *get pissed off* refer to the idea of getting angry not because of conceptual metaphor but because the words *stack, ceiling, lids*, and *pissed* have meanings that at a higher level of generalization refer to the idea of anger. Fortunately, a good deal of empirical work in psycholinguistics has investigated the metaphoric motivation for idiomatic meaning. These psycholinguistic studies employ different methodologies to capture what people ordinarily, and unconsciously, do when they comprehend and make sense of idioms.

One way of uncovering metaphorical knowledge in idiomaticity is through a detailed examination of speakers' mental images for idioms (Gibbs & O'Brien 1990). Consider the idiom *spill the beans*. Try to form a mental image for this phrase and then ask yourself the following questions (Lakoff 1987). Where are the beans before they are spilled? How big is the container? What caused the beans to spill? Is the spilling accidental or intentional? Once they've been

spilled, are the beans in a nice, neat pile? Where are the beans supposed to be? After the beans are spilled, are they easy to retrieve?

Most speakers can form mental images for idioms like *spill the beans* and answer these questions about their mental images without difficulty. Even people without a conscious image for this phrase can answer these questions. Participants in one set of experiments were asked to describe verbally their mental images for idioms with similar figurative meanings in as much detail as possible (Gibbs & O'Brien 1990). We also queried subjects about the causation, intentionality, manner, consequences, and reversibility of the events described in their mental images (What caused the beans to spill? Was the spilling done intentionally or by accident? Where were the beans once they were spilled? Is it easy to get beans back into the container?).

We expected a high degree of consistency in participants' understanding of their mental images for idioms with similar meanings because of the constraints conceptual metaphors (e.g., THE MIND IS A CONTAINER, IDEAS ARE PHYSICAL ENTITIES, and ANGER IS HEAT) impose on the link between idiomatic phrases and their nonliteral meanings. If people's tacit knowledge of idioms is not structured by different conceptual metaphors, there should be little consistency in participants' responses to questions about the causes and consequences of actions within their mental images for idioms with similar nonliteral interpretations.

The data we obtained supported our hypothesis. Participants' mental images for idioms with similar figurative meanings were highly consistent with 75% of their mental images for the different groups of idioms involving similar general images. These general schemata for people's images were not simply representative of the idioms' figurative meanings but captured more specific aspects of the kinesthetic events with the images. For example, idioms such as *flip your lid* and *hit the ceiling* both figuratively mean 'to get angry,' but participants specifically imagined for these phrases some force causing a container to release pressure in a violent manner. There is nothing in the surface forms of these different idioms to constrain tightly the images participants reported. After all, lids can be flipped and ceilings can be hit in a wide variety of ways, due to many different circumstances. But our participants' protocols revealed little variation in the general events that took place in their images for idioms with similar meanings.

Our subjects were also quite consistent in their responses to the different probe questions about their mental images for idioms (over 88%). The probe question data were particularly useful for showing how our understanding of idioms is motivated by different conceptual metaphors. Consider the most frequent responses to the probe questions for the Anger idioms (e.g., *blow your stack, flip your lid, hit the ceiling*). When imagining anger idioms people know that pressure (i.e., stress or frustration) causes the action; that one has little control over the pressure once it builds; that its violent release is unintended (e.g., the blowing of the stack); and that once the release has taken place (i.e., once the ceiling has been hit, the lid flipped, the stack blown), it is difficult to reverse the action. Each of these responses is based on people's folk conceptions of heated fluid or vapor building up and escaping from containers (ones that our participants most frequently reported to be the size of a person's

head). The motivation for these particular folk conceptions comes from two conceptual metaphors—ANGER IS PRESSURIZED HEAT and THE MIND IS A CONTAINER. The mapping of information from different source (e.g., heated fluid in a container) and target (e.g., the anger emotion) domains limits our conceptualization of anger and motivates the idiomatic expressions we use to talk about anger.

These mental imagery studies support the idea that the figurative meanings of idioms are partly motivated by various conceptual metaphors that exist independently as part of our conceptual system. Traditional theories of idiomaticity have no way of accounting for these imagery findings because they assume that the meanings of idioms arise from metaphors that are now dead and no longer a prominent part of our everyday conceptual system. Similarly, more recent linguistic accounts that suppose that idiomatic meanings arise from generalizations across lexical items in these phrases are hard put to account for the mental imagery data. How, for example, do theories based on lexical generalizations account for the specific inferences that people make for anger idioms, for instance, that internal pressure causes the angry event; that the anger action is involuntary; and that the action is performed in a rapid, violent manner? My argument is that lexical theories cannot explain the presence of these specific inference patterns.

Psycholinguistic research has gone on to show that people's knowledge of the metaphorical links between different source and target domains provides the basis for the appropriate use and interpretation of idioms in particular discourse situations (Gibbs & Nayak 1991; Nayak & Gibbs 1990). Participants in one study, for example, gave higher appropriateness ratings to *blew her stack* in a story that described the woman's anger as being like heat in a pressurized container, whereas *bit his head off* was seen as more appropriate in a story that described the woman's anger in terms of a ferocious animal. *Bite your head off* makes sense because people can link the lexical items in this phrase to the conceptual metaphor ANGRY BEHAVIOR IS ANIMAL BEHAVIOR. An animal jumping down a victim's throat is similar to someone shouting angrily. On the other hand, people understand the figurative meaning of *blow your stack* through the conceptual metaphor ANGER IS HEATED FLUID IN A CONTAINER, where a person shouting angrily has the same explosive effect as does the top of a container blowing open under pressure. Thus, readers' judgments about the appropriateness of an idiom in context were influenced by the coherence between the metaphorical information depicted in a text and the conceptual metaphor underlying an idiom's figurative meaning. Even though we may have many idiomatic phrases that refer to a single concept (e.g., anger), some of these phrases may be motivated by different underlying conceptual metaphors (e.g., *blow your stack* vs. *bite your head off*). Because our ordinary concepts are often understood via multiple and sometimes contradictory metaphors, it is no wonder that we have so many different kinds of idioms to reflect the sometimes subtly different aspects of our everyday experience.

One important consequence of the idea that idioms reflect the metaphorical mappings between source and target domains is that idioms are held to have more complex meanings than are their typical literal paraphrases. These idiomatic meanings can be partly predicted, based on the independent assessment

of people's folk understanding of particular source domains that are part of the metaphorical mappings motivating these idioms' interpretations. That is, by looking at the inferences that arise from the mapping of heated fluid in a container onto the idea of anger, one can make specific predictions about what various idioms motivated by ANGER IS HEATED FLUID IN A CONTAINER actually mean.

The results from several experiments explicitly showed that people's understanding of idiomatic meaning reflects the particular entailments of underlying conceptual metaphors (Gibbs 1992). Participants in a first study were questioned about their understanding of events corresponding to particular source domains in various conceptual metaphors (e.g., the source domain of heated fluid in a container for ANGER IS HEATED FLUID IN A CONTAINER). For example, when presented with the scenario of a sealed container filled with fluid, the participants were asked about causation (e.g., What would cause the container to explode?), the intentionality (e.g., Does the container explode on purpose or does it explode through no volition of its own?), and manner (e.g., Does the explosion of the container occur in a gentle or violent manner?).

Overall, the participants in this study were remarkably consistent in their responses to the various questions. To give one example, people responded that the cause of a sealed container exploding out its contents is the internal pressure caused by the increase in the heat of the fluid inside the container, that this explosion is unintentional because containers and fluid have no intentional agency, and that the explosion occurs in a violent manner. More interesting, though, is that people's intuitions about various source domains maps onto their conceptualizations of different target domains in very predictable ways. Thus, other studies in this series showed that when people understand anger idioms such as *blow your stack, flip your lid,* or *hit the ceiling*, they infer that the cause of the anger is internal pressure, that the expression of anger is unintentional, and that the expression occurs in an abrupt and violent manner. However, people do not draw inferences about causation, intentionality, and manner when comprehending literal paraphrases of idioms, such as *get very angry*. Literal phrases, such as *get very angry*, are not motivated by the same set of conceptual metaphors as are specific idioms such as *blow your stack*. For this reason, people do not view the meanings of *blow your stack* and *get very angry* as equivalent, despite their apparent similarity.

A final series of reading-time experiments showed that people find idioms more appropriate and easier to understand when they are seen in discourse contexts that are consistent with the various entailments of these phrases. Control studies showed that these differences in the interpretation of idioms and their literal paraphrases cannot be attributed to differences in the entailments of their respective verbs and nouns. Thus, the meanings of the individual words in idioms are not by themselves sufficient to account for the complex inferences people make about the meanings of idioms.

This set of studies on the conceptual basis of idiomatic meaning provides experimental evidence in support of cognitive linguistic analyses of idiomaticity. Such data specifically support the idea that the mappings of source-to-target domain information in conceptual metaphors preserve the structural characteristics or cognitive typology of the source domains (Lakoff 1990). The

data from these studies are important because they provide an independent, nonlinguistic way of partially predicting what specific meanings some idioms are likely to possess, based on the analyses of certain metaphorical concepts in long-term memory. As such, this experimental work is an important, perhaps necessary, complement to cognitive linguistic analyses of idiomaticity.

Metaphor and Immediate Idiom Comprehension

The experimental evidence on the conceptual basis for interpreting idioms is not representative of evidence used in contemporary psycholinguistic research. Psycholinguists traditionally attempt to formulate theories of linguistic understanding to account for moment-by-moment language processing. Only experimental methodologies that tap into what people actually, and unconsciously, do online at the very moment when comprehension occurs are thought to be appropriate in studying normal utterance understanding. People might tacitly recognize that idioms have meanings that are motivated by different kinds of conceptual knowledge. But this does not mean that people access this conceptual knowledge each and every time they encounter certain idioms.

Several sets of studies are currently being conducted to examine whether people actually access metaphorical knowledge during the immediate, online processing of idioms. Participants in a first study read simple stories one line at a time on a computer screen with each story ending in one of three different phrases. The following is an example of one story along with each of its different final phrases.

(2) John lent his new car to a friend, Sally.
When Sally later returned the car, the front end was badly damaged.
When Sally showed John the car,
He blew his stack. (*appropriate idiom*)
He got very angry. (*literal paraphrase*)
He saw many dents. (*control phrase*)

After reading the final phrase and pushing the comprehension button, the participants were immediately presented with a letter string on the computer screen and their task was to decide as quickly as possible if the letter string constituted an English word (i.e., a lexical decision task). These letter strings or targets were either words that represented a conceptual metaphor motivating the appropriate idiom (e.g., *heat*, which represents ANGER IS HEATED FLUID IN A CONTAINER), or nonword letter strings (e.g., *saet*).

If people actually access specific conceptual metaphors (e.g., ANGER IS HEATED FLUID) while comprehending certain idioms phrases (e.g., *He blew his stack*), then this activated metaphorical knowledge should facilitate or prime participants' responses to the metaphor targets (e.g., *heat*), such that they respond more quickly than after reading either the literal paraphrases or the control phrases. In fact, participants responded significantly faster to the metaphoric targets when they had just read the idioms rather than either the literal phrases or the control expressions. Of course, people may respond faster to *heat* after having read *blew his stack* simply because of some preexisting semantic associations between the literal words in the expressions. Yet a follow-up study

showed that people were not faster in responding to *heat* when they read *blow the stack* in a literal context.

These findings provide some initial evidence that people normally access the underlying conceptual metaphors for idioms when they process these phrases. Follow-up experiments will investigate other aspects of the hypothesis that people automatically access metaphorical knowledge during their immediate, online processing of idioms. One important aim of this work is to acknowledge explicitly that linguistic understanding encompasses many different levels ranging from quick, unconscious mental processes up to more conscious, reflective analysis of meaning (Gibbs 1994, 1996). Each of these different kinds of understanding requires different methodological tools to fully analyze it, and this is why the study of idiomaticity demands the expertise of linguists, psychologists, and computer scientists. We should be careful to recognize that how people make sense of why idioms mean what they do provides different insights into idiom interpretation than do studies on the immediate comprehension of idioms. For instance, it might very well be the case that with additional research, little evidence will support the idea that people automatically access conceptual metaphors during idiom understanding (see Glucksberg, Brown, & McGlone 1993) even though there are plenty of data to suggest that people have some sense of why idioms mean what they do because of underlying conceptual metaphors. The conceptual view of idiomaticity emphasized here does not necessarily predict that conceptual metaphors influence all aspects of idiom understanding. As with many issues, the ultimate answer to the question of conceptual metaphors' role in idiom interpretation will be an empirical matter.

Idioms and Metonymy

Most of the focus on the conceptual basis of idiomaticity has been on the role that metaphor plays in motivating what idioms mean figuratively. However, other figurative schemes of thought also give rise to different idioms and help motivate idiom meaning for contemporary speakers. For example, metonymy is a fundamental part of our conceptual system whereby people take one well-understood or easily perceived aspect of something to represent or stand for the thing as a whole.

Various metonymic models in our conceptual system underlie the use of many kinds of figurative and conventional expressions. As with metaphor, this is best illustrated by considering the similarity between different metonymic expressions that reflect, for instance, the metonymic mappings of OBJECT USED FOR USER (e.g., *The buses are on strike, Our sax has the flu today, We need a new glove at third base*), CONTROLLER FOR CONTROLLED (e.g., *Nixon bombed Hanoi, A BMW rear-ended me, Napoleon lost at Waterloo*), and THE PLACE FOR THE EVENT (e.g., *Wall Street is in a panic, Hollywood is putting out terrible movies this year, Paris has dropped hemlines this year*).

Many thousands of idioms reflect metonymic modes of thought. Each of these expressions reflects some salient aspect of an object, idea, or event and then stands for or represents the object, idea, or event as a whole. For instance, *bite the dust* (meaning 'to die'), *throw in the towel* (meaning 'to give up on some

activity'), and *pass the buck* (meaning 'to ignore one's responsibility') all refer to salient acts in series of events. Even the classical phrase *kick the bucket* is metonymic in referring to the last live act that a pig does before dying. In each case, a salient act has a "stands-for" relationship to an entire idea or event.

Although metaphor and metonymy individually motivate different kinds of linguistic expressions, there are many cases where these tropes are combined in idiomatic language. Consider first how we get metaphors for which there is a link with their metonymic origins (Goossens 1990). One instance of this is the phrase *to be close-lipped*, meaning 'to be silent or to say little.' *Close-lipped* can be literally paraphrased as 'having the lips close together' or as 'having the lips closed.' When *close-lipped* is used to indicate that a person is literally silent, we therefore need the metonymic reading. If, on the other hand, we describe as *close-lipped* someone who is actually talking a lot but does not give away what one would really want to hear from him, we have a metaphor (given the saliency of the metonymic reading, we have a metaphor from metonymy).

Another kind of interaction is metonymy within metaphor. Consider the phrase *shoot your mouth off* 'to talk foolishly about something that one does not know much about or should not talk about.' The source domain in this metaphorical mapping is the foolish use of firearms that is mapped onto the target domain of thoughtless linguistic action. When the word *mouth* is integrated into a scene relating to the use of firearms, it must be reinterpreted as having the properties of the gun alluded to in the phrase *shoot your mouth off*. In the target domain, however, there is a first level of interpretation that amounts to something like 'to use your mouth foolishly,' in which *mouth* metonymically stands for the speech faculty. This interaction of metonymy with metaphor explains why *Don't shoot your mouth off* means 'Don't say anything rash.' A similar type of analysis can be applied to other expressions regarding linguistic action, such as *catch someone's eye*.

These analyses of the interaction of metaphor and metonymy in idiomatic expressions for linguistic action illustrate how tropes are frequently combined in different idiomatic expressions. I see this issue of trope interaction in idioms to be one of the exciting avenues for future research.

Conclusion

In this chapter I have argued that we must recognize that many idioms are analyzable with their components independently contributing to what these phrases mean figuratively. The acquisition and comprehension of idioms are based on compositional parsing strategies that are similar to those employed in the comprehension of literal speech. Listeners and readers do not switch from a literal to a nonliteral mode of processing when comprehending idioms. Instead, they rely on a fast, unconscious process whereby they seek to discover the independent meanings of the parts of idioms and combine these to recognize what idioms mean as wholes. These meanings are not necessarily the literal meanings of the words in idioms but, instead, may merely reflect figurative interpretations of different words and word combinations in context. Our theoretical understanding of the idiom comprehension process will be limited until we develop more sophisticated accounts of lexical semantics.

My main aim in this chapter has been to demonstrate how people's preexisting metaphorical understanding of many basic concepts partly motivates what people see as the figurative meanings of idioms and their components. This metaphorical knowledge influences the linguistic behavior of idioms as well as the learning, understanding, and real-time processing of idioms. Idiom researchers must be able to incorporate these data in their accounts of different aspects of idiomaticity. Idiom researchers must also begin to explicitly recognize the difference between dead metaphors and conventional metaphors that reflect ordinary patterns of human conceptual structure. These conventional metaphoric, and to some extent metonymic, mappings not only motivate the meanings of idiomatic phrases, but also explain our creation and use of many conventional and, to some observers, literal expressions, such as *I exploded with anger*, which is motivated by the same metaphor of ANGER IS HEATED FLUID IN A CONTAINER as is *blow your stack, flip your lid,* and *hit the ceiling*. These same underlying conceptual metaphors are also elaborated on in novel linguistic metaphors such as those found in poetry (Gibbs & Nascimento 1994; Lakoff & Turner 1989). In general, then, idiomatic phrases do not arise from some unique linguistic and conceptual knowledge; instead they reflect ordinary, figurative patterns of human understanding.

Several linguists and psychologists have argued with the foregoing conclusions about the conceptual basis of idiomaticity. Psycholinguists have suggested that idioms do not reflect much about human conceptual structure (Keysar & Bly 1999; Ortony 1988) or, at the very least, are not used in people's ordinary comprehension of idiomatic phrases (Glucksberg, Brown, & McGlone 1993). Several scholars have argued that the conceptual view of idiomaticity advocated here fails to acknowledge the importance of lexical information in idiomatic meaning (Kreuz & Graesser 1991; Stock, Slack, & Ortony 1993; and see Gibbs & Nayak 1991).

Yet the conceptual view claims only that conceptual metaphors provide part of the link between idiomatic phrases and their overall figurative interpretations. The meanings of an idiom's components (and again, what these meanings are needs to be better understood) contribute significantly to idiom meanings. But lexical meanings do not by themselves capture the complex inferences associated with idiomatic meanings; this is one reason why conceptual metaphors are an essential part of a theory of idiomaticity. Such a theory can certainly provide some motivated reasons that link idioms to their figurative interpretations. These motivations are directly tied to the ways that people ordinarily think about the concepts to which idioms refer and suggest that people actually conceptualize much of their everyday experience in figurative terms. We should recognize that many idioms are partly motivated by figures of thought that make up a significant part of our ordinary conceptual structures.

References

Bolinger, D. (1971) *The Phrasal Verb in English*, Harvard University Press, Cambridge, Massachusetts.

Cacciari, C. (1993) "The Place of Idioms in a Literal and Metaphorical World," in C. Cacciari and P. Tabossi, eds., *Idioms: Processing, Structure, and Interpretation*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Cacciari, C., and S. Glucksberg (1990) "Understanding Idiomatic Expressions: The Contribution of Word Meanings," in G. Simpson, ed., *Understanding Word and Sentence,* Elsevier, Amsterdam.

Cacciari, C., and M. Levorato (1989) "How Children Understand Idioms in Discourse," *Journal of Child Language* 16, 387–405.

Cacciari, C., and P. Tabossi (1988) "The Comprehension of Idioms," *Journal of Memory and Language* 27, 668–683.

Cacciari, C., and P. Tabossi, eds. (1993) *Idioms: Processing. Structure, and Interpretation*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Chafe, W. (1970) *Meaning and the Structure of Language*, University of Chicago Press, Chicago.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*, MIT Press, Cambridge, Massachusetts.

Chomsky, N. (1980) *Rules and Representations*, Columbia University Press, New York.

Coulmas, F. (1981) "Idiomaticity as a Problem of Pragmatics," in H. Parret and M. Sbisa, eds., *Possibilities and Limitations of Pragmatics*, Lawrence Erlbaum Associates, Amsterdam.

Fillmore, C., P. Kay, and M. O'Connor (1988) "Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone," Language* 64, 501–538.

Flores d'Arcais, G. (1993) "The Comprehension and Semantic Interpretation of Idioms," in C. Cacciari and P. Tabossi, eds., *Idioms: Processing, Structure, and Interpretation*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Fraser, B. (1970) "Idioms Within a Transformational Grammar," *Foundations of Language* 6, 22–42.

Gibbs, R. (1980) "Spilling the Beans on Understanding and Memory for Idioms in Conversation," *Memory and Cognition* 8, 449–456.

Gibbs, R. (1984) "Literal Meaning and Psychological Theory," *Cognitive Science* 8, 275–304.

Gibbs, R. (1985) "On the Process of Understanding Idioms," *Journal of Psycholinguistic Research* 14, 465–472.

Gibbs, R. (1986) "Skating on Thin Ice: Literal Meaning and Understanding Idioms in Conversation," *Discourse Processes* 9, 17–30.

Gibbs, R. (1987) "Linguistic Factors in Children's Understanding of Idioms," *Journal of Child Language* 14, 569–586.

Gibbs, R. (1989) "Understanding and Literal Meaning," *Cognitive Science* 13, 243–251.

Gibbs, R. (1991) "Semantic Analyzability in Children's Understanding of Idioms," *Journal of Speech and Hearing Research* 34, 613–620.

Gibbs, R. (1992) "What Do Idioms Really Mean?" *Journal of Memory and Language* 31, 385–406.

Gibbs, R. (1993) "Why Idioms Are Not Dead Metaphors," in C. Cacciari and P. Tabossi, eds., *Idioms: Processing, Structure, and Interpretation*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Gibbs, R. (1994) *The Poetics of Mind: Figurative Thought, Language, and Understanding*, Cambridge University Press, New York.

Gibbs, R. (1996) "What's Cognitive about Cognitive Linguistics?" in G. Casad, ed., *Cognitive Linguistics in the Redwoods*, Berlin: Mouton de Gruyter.

Gibbs, R., D. Buchalter, J. Moise, and W. Farrar (1993) "Literal Meaning and Figurative Language," *Discourse Processes* 16, 387–404.

Gibbs, R. and S. Nascimento (1994) "How We Talk When We Talk About Love: Metaphorical Concepts and Understanding Love Poetry," in R. Kreuz and M. MacNealy, eds., *Empirical and Aesthetic Approaches to Literature*, Ablex, Norwood, New Jersey.

Gibbs, R. and N. Nayak (1989) "Psycholinguistic Studies on the Syntactic Behavior of Idioms," *Cognitive Psychology* 21, 100–138.

Gibbs, R. and N. Nayak (1991) "Why Idioms Mean What They Do," *Journal of Experimental Psychology: General* 120, 93–95.

Gibbs, R., N. Nayak, J. Bolton, and M. Keppel (1989) "Speakers' Assumptions About the Lexical Flexibility of Idioms," *Memory & Cognition* 17, 58–68.

Gibbs, R., N. Nayak, and C. Cutting (1989) "How To Kick the Bucket and Not Decompose: Analyzability and Idiom Processing," *Journal of Memory and Language* 28, 576–593.

Gibbs, R. and J. O'Brien (1990) "Idioms and Mental Imagery: The Metaphorical Motivation for Idiomatic Meaning," *Cognition* 36, 35–68.

Glucksberg, S. (1993) "Idiom Meanings and Allusional Content," in C. Cacciari and P. Tabossi, eds., *Idioms: Processing, Structure, Interpretation*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Glucksberg, S., M. Brown, and M. McGlone (1993) "Conceptual Metaphors Are Not Automatically Accessed During Idiom Comprehension," *Memory & Cognition* 21, 711–719.

Goossens, L. (1990) "Metaphtonymy: The Interaction of Metaphor and Metonymy in Expressions for Linguistic Action," *Cognitive Linguistics* 1, 323–340.

Johnson, M. (1987) *The Body in the Mind*, University of Chicago Press, Chicago.
Katz, J. (1973) "Compositionality, Idiomaticity, and Lexical Substitution," in S. Anderson and P. Kiparsky, eds., *A Festschrift for Morris Halle*, Holt, Rinehart & Winston, New York.
Keysar, B. and B. Bly (1999) "Swimming Against the Current: Do Idioms Reflect Conceptual Structure?" *Journal of Progmatics* 31, 1559–1578.
Kovecses, Z. (1986) *Metaphors of Anger, Pride, and Love*, John Benjamins, Amsterdam.
Kreuz, R. and A. Graesser (1991) "Aspects of Idiom Comprehension: Comment on Nayak and Gibbs," *Journal of Experimental Psychology: General* 120, 90–92.
Lakoff, G. (1986) "The Meanings of Literal," *Metaphor and Symbolic Activity* 1, 291–296.
Lakoff, G. (1987) "*Women, Fire, and Dangerous Things*," Chicago University Press, Chicago.
Lakoff, G. (1990) "The Invariance Hypothesis: Is Abstract Reason Based on Image-Schemas?" *Cognitive Linguistics* 1, 39–74.
Lakoff, G. and M. Johnson (1980) *Metaphors We Live By*, Chicago University Press, Chicago.
Lakoff, G. and M. Turner (1989) *More Than Cool Reason: A Field Guide to Poetic Metaphor*, University of Chicago Press, Chicago.
Lambrecht, K. (1984) "Formulaicity, Frame Semantics, and Pragmatics in German Binomial Expressions," *Language* 60, 753–796.
Langacker, R. (1986) *Foundations of Cognitive Grammar* (Vol. 1), Stanford University Press, Stanford, California.
Levorato, M. (1993) "The Acquisition of Idioms and the Development of Figurative Competence," in C. Cacciari and P. Tabossi, eds., *Idioms: Processing, Structure, and Interpretation*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.
Levorato, M. and C. Cacciari (1992) "Children's Comprehension and Production of Idioms: The Role of Context and Familiarity," *Journal of Child Language* 19, 415–433.
Lindner, S. (1981) A Lexico-Semantic Analysis of Verb-Particle Constructions With *Up* and *Out*, Doctoral dissertation, University of California, San Diego.
McGlone, M., S. Glucksberg, and C. Cacciari (1994) "Semantic Productivity and Idiom Comprehension." *Discourse Processes* 17(2), 167–190.
Munro, P. (1989) *Slang U: The Official Dictionary of College Slang*, Harmony Books, New York.
Nayak, N. and R. Gibbs (1990) "Conceptual Knowledge in the Interpretation of Idioms," *Journal of Experimental Psychology: General* 119, 315–330.
Nippold, M. and S. Martin (1989) "Idiom Interpretation in Isolation Versus Context: A Developmental Study with Adolescents," *Journal of Speech and Hearing Research* 32, 59–66.
Nunberg, G. (1978) *The Pragmatics of Reference*, Indiana University Linguistics Club, Bloomington.
Ortony, A. (1988) "Are Emotion Metaphors Conceptual or Lexical?" *Cognition and Emotion* 2, 95–103.
Peterson, R. and C. Burgess (1993) "Syntactic and Semantic Processing During Idiom Comprehension: Neurolinguistic and Psycholinguistic Dissociations," in C. Cacciari and P. Tabossi, eds., *Idioms: Processing, Structure, and Interpretation*, Lawrence Erilbaum Associates, Hillsdale, New Jersey.
Ruwet, N. (1992) *Syntax and Human Experience*, University of Chicago Press, Chicago.
Searle, J. (1979) "Literal Meaning," in J. Searle, ed., *Expression and Meaning*, Cambridge University Press, New York.
Stock, O., J. Slack, and A. Ortony (1993) "Building Castles in the Air: Some Computational and Theoretical Issues in Idiom Comprehension," in C. Cacciari and P. Tabossi, eds., *Idioms: Processing, Structure, and Interpretation*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.
Sweetser, E. (1990) *From Etymology to Pragmatics: The Mind–Body Metaphor in Semantic Structure and Semantic Change*, Cambridge University Press, New York.
Tabossi, P. and F. Zardon (1993) "The Activation of Idiomatic Meaning in Spoken Language Comprehension," in C. Cacciari and P. Tabossi, eds., *Idioms: Processing, Structure, and Interpretation*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.
Turner, M. (1991) *Reading Minds: The Study of English in the Age of Cognitive Science*, Princeton University Press, Princeton, New Jersey.
Wasow, T., I. Sag, and G. Nunberg (1983) "Idioms: An Interim Report," in S. Hattori and K. Inoue, eds., *Proceedings of the XIIIth International Congress of Linguists*, CIPL, Tokyo.
Weinreich, U. (1969) "Problems in the Analysis of Idioms," in J. Puhvel, ed., *Substance and Structure of Language*, University of California Press, Los Angeles.

PART XVII

Intelligence

Chapter 34

In a Nutshell

Howard Gardner

Allow me to transport all of us to the Paris of 1900—La Belle Epoque—when the city fathers of Paris approached a psychologist named Alfred Binet with an unusual request: Could he devise some kind of a measure that would predict which youngsters would succeed and which would fail in the primary grades of Paris schools? As everybody knows, Binet succeeded. In short order, his discovery came to be called the "intelligence test"; his measure, the "IQ." Like other Parisian fashions, the IQ soon made its way to the United States, where it enjoyed a modest success until World War I. Then, it was used to test over one million American recruits, and it had truly arrived. From that day on, the IQ test has looked like psychology's biggest success—a genuinely useful scientific tool.

What is the vision that led to the excitement about IQ? At least in the West, people had always relied on intuitive assessments of how smart other people were. Now intelligence seemed to be quantifiable. You could measure someone's actual or potential height, and now, it seemed, you could also measure someone's actual or potential intelligence. We had one dimension of mental ability along which we could array everyone.

The search for the perfect measure of intelligence has proceeded apace. Here, for example, are some quotations from an ad for a widely used test:

> Need an individual test which quickly provides a stable and reliable estimate of intelligence in four or five minutes per form? Has three forms? Does not depend on verbal production or subjective scoring? Can be used with the severely physically handicapped (even paralyzed) if they can signal yes or no? Handles two-year-olds and superior adults with the same short series of items and the same format? Only $16.00 complete.

Now, that's quite a claim. The American psychologist Arthur Jensen suggests that we could look at reaction time to assess intelligence: a set of lights go on; how quickly can the subject react? The British psychologist Hans Eysenck suggests that investigators of intelligence should look directly at brain waves.

There are also, of course, more sophisticated versions of the IQ test. One of them is called the Scholastic Aptitude Test (SAT). It purports to be a similar kind of measure, and if you add up a person's verbal and math scores, as is often done, you can rate him or her along a single intellectual dimension.

From chapter 1 in *Frames of Mind: The Theory of Multiple Intelligences* (New York: Basic Books, 1993), 5–12. Reprinted with permission.

Programs for the gifted, for example, often use that kind of measure; if your IQ is in excess of 130, you're admitted to the program.

I want to suggest that along with this one-dimensional view of how to assess people's minds comes a corresponding view of school, which I will call the "uniform view." In the uniform school, there is a core curriculum, a set of facts that everybody should know, and very few electives. The better students, perhaps those with higher IQs, are allowed to take courses that call upon critical reading, calculation, and thinking skills. In the "uniform school," there are regular assessments, using paper and pencil instruments, of the IQ or SAT variety. They yield reliable rankings of people; the best and the brightest get into the better colleges, and perhaps—but only perhaps—they will also get better rankings in life. There is no question but that this approach works well for certain people—schools such as Harvard are eloquent testimony to that. Since this measurement and selection system is clearly meritocratic in certain respects, it has something to recommend it.

But there is an alternative vision that I would like to present—one based on a radically different view of the mind, and one that yields a very different view of school. It is a pluralistic view of mind, recognizing many different and discrete facets of cognition, acknowledging that people have different cognitive strengths and contrasting cognitive styles. I would also like to introduce the concept of an individual-centered school that takes this multifaceted view of intelligence seriously. This model for a school is based in part on findings from sciences that did not even exist in Binet's time: cognitive science (the study of the mind), and neuroscience (the study of the brain). One such approach I have called my "theory of multiple intelligences." Let me tell you something about its sources, its claims, and its educational implications for a possible school of the future.

Dissatisfaction with the concept of IQ and with unitary views of intelligence is fairly widespread—one thinks, for instance, of the work of L. L. Thurstone, J. P. Guilford, and other critics. From my point of view, however, these criticisms do not suffice. The whole concept has to be challenged; in fact, it has to be replaced.

I believe that we should get away altogether from tests and correlations among tests, and look instead at more naturalistic sources of information about how peoples around the world develop skills important to their way of life. Think, for example, of sailors in the South Seas, who find their way around hundreds, or even thousands, of islands by looking at the constellations of stars in the sky, feeling the way a boat passes over the water, and noticing a few scattered landmarks. A word for intelligence in a society of these sailors would probably refer to that kind of navigational ability. Think of surgeons and engineers, hunters and fishermen, dancers and choreographers, athletes and athletic coaches, tribal chiefs and sorcerers. All of these different roles need to be taken into account if we accept the way I define intelligence—that is, as the ability to solve problems, or to fashion products, that are valued in one or more cultural or community settings. For the moment I am saying nothing about whether there is one dimension, or more than one dimension, of intelligence; nothing about whether intelligence is inborn or developed. Instead I emphasize the ability to solve problems and to fashion products. In my work I seek the

building blocks of the intelligences used by the aforementioned sailors and surgeons and sorcerers.

The science in this enterprise, to the extent that it exists, involves trying to discover the *right* description of the intelligences. What is an intelligence? To try to answer this question, I have, with my colleagues, surveyed a wide set of sources which, to my knowledge, have never been considered together before. One source is what we already know concerning the development of different kinds of skills in normal children. Another source, and a very important one, is information on the ways that these abilities break down under conditions of brain damage. When one suffers a stroke or some other kind of brain damage, various abilities can be destroyed, or spared, in isolation from other abilities. This research with brain-damaged patients yields a very powerful kind of evidence, because it seems to reflect the way the nervous system has evolved over the millennia to yield certain discrete kinds of intelligence.

My research group looks at other special populations as well: prodigies, idiot savants, autistic children, children with learning disabilities, all of whom exhibit very jagged cognitive profiles—profiles that are extremely difficult to explain in terms of a unitary view of intelligence. We examine cognition in diverse animal species and in dramatically different cultures. Finally, we consider two kinds of psychological evidence: correlations among psychological tests of the sort yielded by a careful statistical analysis of a test battery; and the results of efforts of skill training. When you train a person in skill A, for example, does that training transfer to skill B? So, for example, does training in mathematics enhance one's musical abilities, or vice versa?

Obviously, through looking at all these sources—information on development, on breakdowns, on special populations, and the like—we end up with a cornucopia of information. Optimally, we would perform a statistical factor analysis, feeding all the data into a computer and noting the kinds of factors or intelligences that are extracted. Alas, the kind of material with which I was working didn't exist in a form that is susceptible to computation, and so we had to perform a more subjective factor analysis. In truth, we simply studied the results as best we could, and tried to organize them in a way that made sense to us, and hopefully, to critical readers as well. My resulting list of seven intelligences is a preliminary attempt to organize this mass of information.

I want now to mention briefly the seven intelligences we have located, and to cite one or two examples of each intelligence. Linguistic intelligence is the kind of ability exhibited in its fullest form, perhaps, by poets. Logical-mathematical intelligence, as the name implies, is logical and mathematical ability, as well as scientific ability. Jean Piaget, the great developmental psychologist, thought he was studying *all* intelligence, but I believe he was studying the development of logical-mathematical intelligence. Although I name the linguistic and logical-mathematical intelligences first, it is not because I think they are the most important—in fact, I am convinced that all seven of the intelligences have equal claim to priority. In our society, however, we have put linguistic and logical-mathematical intelligences, figuratively speaking, on a pedestal. Much of our testing is based on this high valuation of verbal and mathematical skills. If you do well in language and logic, you should do well in IQ tests and SATs, and you may well get into a prestigious college, but whether you do well once you

leave is probably going to depend as much on the extent to which you possess and use the other intelligences, and it is to those that I want to give equal attention.

Spatial intelligence is the ability to form a mental model of a spatial world and to be able to maneuver and operate using that model. Sailors, engineers, surgeons, sculptors, and painters, to name just a few examples, all have highly developed spatial intelligence. Musical intelligence is the fourth category of ability we have identified: Leonard Bernstein had lots of it; Mozart, presumably, had even more. Bodily-kinesthetic intelligence is the ability to solve problems or to fashion products using one's whole body, or parts of the body. Dancers, athletes, surgeons, and craftspeople all exhibit highly developed bodily-kinesthetic intelligence.

Finally, I propose two forms of personal intelligence—not well understood, elusive to study, but immensely important. Interpersonal intelligence is the ability to understand other people: what motivates them, how they work, how to work cooperatively with them. Successful salespeople, politicians, teachers, clinicians, and religious leaders are all likely to be individuals with high degrees of interpersonal intelligence. Intrapersonal intelligence, a seventh kind of intelligence, is a correlative ability, turned inward. It is a capacity to form an accurate, veridical model of oneself and to be able to use that model to operate effectively in life.

These, then, are the seven intelligences that we have uncovered and described in our research. This is a preliminary list, as I have said; obviously, each form of intelligence can be subdivided, or the list can be rearranged. The real point here is to make the case for the plurality of intellect. Also, we believe that individuals may differ in the particular intelligence profiles with which they are born, and that certainly they differ in the profiles they end up with. I think of the intelligences as raw, biological potentials, which can be seen in pure form only in individuals who are, in the technical sense, freaks. In almost everybody else the intelligences work together to solve problems, to yield various kinds of cultural endstates—vocations, avocations, and the like.

This is my theory of multiple intelligence in capsule form. In my view, the purpose of school should be to develop intelligences and to help people reach vocational and avocational goals that are appropriate to their particular spectrum of intelligences. People who are helped to do so, I believe, feel more engaged and competent, and therefore more inclined to serve the society in a constructive way.

These thoughts, and the critique of a universalistic view of mind with which I began, lead to the notion of an individual-centered school, one geared to optimal understanding and development of each student's cognitive profile. This vision stands in direct contrast to that of the uniform school that I described earlier.

The design of my ideal school of the future is based upon two assumptions. The first is that not all people have the same interests and abilities; not all of us learn in the same way. (And we now have the tools to begin to address these individual differences in school.) The second assumption is one that hurts: it is the assumption that nowadays no one person can learn everything there is to learn. We would all like, as Renaissance men and women, to know everything,

or at least to believe in the potential of knowing everything, but that ideal clearly is not possible anymore. Choice is therefore inevitable, and one of the things that I want to argue is that the choices that we make for ourselves, and for the people who are under our charge, might as well be informed choices. An individual-centered school would be rich in assessment of individual abilities and proclivities. It would seek to match individuals not only to curricular areas, but also to particular ways of teaching those subjects. And after the first few grades, the school would also seek to match individuals with the various kinds of life and work options that are available in their culture.

I want to propose a new set of roles for educators that might make this vision a reality. First of all, we might have what I will call "assessment specialists." The job of these people would be to try to understand as sensitively and comprehensively as possible the abilities and interests of the students in a school. It would be very important, however, that the assessment specialists use "intelligence-fair" instruments. We want to be able to look specifically and directly at spatial abilities, at personal abilities, and the like, and not through the usual lenses of the linguistic and logical-mathematical intelligences. Up until now nearly all assessment has depended indirectly on measurement of those abilities; if students are not strong in those two areas, their abilities in other areas may be obscured. Once we begin to try to assess other kinds of intelligences directly, I am confident that particular students will reveal strengths in quite different areas, and the notion of general brightness will disappear or become greatly attenuated.

In addition to the assessment specialist, the school of the future might have the "student-curriculum broker." It would be his or her job to help match students' profiles, goals, and interests to particular curricula and to particular styles of learning. Incidentally, I think that the new interactive technologies offer considerable promise in this area: it will probably be much easier in the future for "brokers" to match individual students to ways of learning that prove comfortable for them.

There should also be, I think, a "school-community broker," who would match students to learning opportunities in the wider community. It would be this person's job to find situations in the community, particularly options not available in the school, for children who exhibit unusual cognitive profiles. I have in mind apprenticeships, mentorships, internships in organizations, "big brothers," "big sisters"—individuals and organizations with whom these students might work to secure a feeling for different kinds of vocational and avocational roles in the society. I am not worried about those occasional youngsters who are good in everything. They're going to do just fine. I'm concerned about those who don't shine in the standardized tests, and who, therefore, tend to be written off as not having gifts of any kind. It seems to me that the school-community broker could spot these youngsters and find placements in the community that provide chances for them to shine.

There is ample room in this vision for teachers, as well, and also for master teachers. In my view, teachers would be freed to do what they are supposed to do, which is to teach their subject matter, in their preferred style of teaching. The job of master teacher would be very demanding. It would involve, first of all, supervising the novice teachers and guiding them; but the master teacher

would also seek to ensure that the complex student-assessment-curriculum-community equation is balanced appropriately. If the equation is seriously imbalanced, master teachers would intervene and suggest ways to make things better.

Clearly, what I am describing is a tall order; it might even be called utopian. And there is a major risk to this program, of which I am well aware. That is the risk of premature billeting—of saying, "Well, Johnny is four, he seems to be musical, so we are going to send him to Juilliard and drop everything else." There is, however, nothing inherent in the approach that I have described that demands this early overdetermination—quite the contrary. It seems to me that early identification of strengths can be very helpful in indicating what kinds of experiences children might profit from; but early identification of weaknesses can be equally important. If a weakness is identified early, there is a chance to attend to it before it is too late, and to come up with alternative ways of teaching or of covering an important skill area.

We now have the technological and the human resources to implement such an individual-centered school. Achieving it is a question of will, including the will to withstand the current enormous pressures toward uniformity and unidimensional assessments. There are strong pressures now, which you read about every day in the newspapers, to compare students, to compare teachers, states, even entire countries, using one dimension or criterion, a kind of a crypto-IQ assessment. Clearly, everything I have described today stands in direct opposition to that particular view of the world. Indeed that is my intent—to provide a ringing indictment of such one-track thinking.

I believe that in our society we suffer from three biases, which I have nicknamed "Westist," "Testist," and "Bestist." "Westist" involves putting certain Western cultural values, which date back to Socrates, on a pedestal. Logical thinking, for example, is important; rationality is important; but they are not the only virtues. "Testist" suggests a bias towards focusing upon those human abilities or approaches that are readily testable. If it can't be tested, it sometimes seems, it is not worth paying attention to. My feeling is that assessment can be much broader, much more humane than it is now, and that psychologists should spend less time ranking people and more time trying to help them.

"Bestist" is a not very veiled reference to a book by David Halberstam called *The best and the brightest*. Halberstam referred ironically to figures such as Harvard faculty members who were brought to Washington to help President John F. Kennedy and in the process launched the Vietnam War. I think that any belief that all the answers to a given problem lie in one certain approach, such as logical-mathematical thinking, can be very dangerous. Current views of intellect need to be leavened with other more comprehensive points of view.

It is of the utmost importance that we recognize and nurture all of the varied human intelligences, and all of the combinations of intelligences. We are all so different largely because we all have different combinations of intelligences. If we recognize this, I think we will have at least a better chance of dealing appropriately with the many problems that we face in the world. If we can mobilize the spectrum of human abilities, not only will people feel better about themselves and more competent; it is even possible that they will also feel more engaged and better able to join the rest of the world community in working for

the broader good. Perhaps if we can mobilize the full range of human intelligences and ally them to an ethical sense, we can help to increase the likelihood of our survival on this planet, and perhaps even contribute to our thriving.

Acknowledgments

This chapter is based on an informal talk given at the 350th anniversary of Harvard University on 5 September 1986. The work reported in this article was supported by the Rockefeller Foundation, the Spencer Foundation, and the Bernard Van Leer Foundation.

Chapter 35

A Rounded Version

Howard Gardner and Joseph Walters

Two eleven-year-old children are taking a test of "intelligence." They sit at their desks laboring over the meanings of different words, the interpretation of graphs, and the solutions to arithmetic problems. They record their answers by filling in small circles on a single piece of paper. Later these completed answer sheets are scored objectively: the number of right answers is converted into a standardized score that compares the individual child with a population of children of similar age.

The teachers of these children review the different scores. They notice that one of the children has performed at a superior level; on all sections of the test, she answered more questions correctly than did her peers. In fact, her score is similar to that of children three to four years older. The other child's performance is average—his scores reflect those of other children his age.

A subtle change in expectations surrounds the review of these test scores. Teachers begin to expect the first child to do quite well during her formal schooling, whereas the second should have only moderate success. Indeed these predictions come true. In other words, the test taken by the eleven-year-olds serves as a reliable predictor of their later performance in school.

How does this happen? One explanation involves our free use of the word "intelligence": the child with the greater "intelligence" has the ability to solve problems, to find the answers to specific questions, and to learn new material quickly and efficiently. These skills in turn play a central role in school success. In this view, "intelligence" is a singular faculty that is brought to bear in any problem-solving situation. Since schooling deals largely with solving problems of various sorts, predicting this capacity in young children predicts their future success in school.

"Intelligence," from this point of view, is a general ability that is found in varying degrees in all individuals. It is the key to success in solving problems. This ability can be measured reliably with standardized pencil-and-paper tests that, in turn, predict future success in school.

What happens after school is completed? Consider the two individuals in the example. Looking further down the road, we find that the "average" student has become a highly successful mechanical engineer who has risen to a position of prominence in both the professional community of engineers as well as in civic groups in his community. His success is no fluke—he is considered by all to be a talented individual. The "superior" student, on the other hand, has had

From chapter 2 in *Frames of Mind: The Theory of Multiple Intelligences* (New York: Basic Books, 1993), 13–34. Reprinted with permission.

little success in her chosen career as a writer; after repeated rejections by publishers, she has taken up a middle management position in a bank. While certainly not a "failure," she is considered by her peers to be quite "ordinary" in her adult accomplishments. So what happened?

This fabricated example is based on the facts of intelligence testing. IQ tests predict school performance with considerable accuracy, but they are only an indifferent predictor of performance in a profession after formal schooling (Jencks, 1972). Furthermore, even as IQ tests measure only logical or logical-linguistic capacities, in this society we are nearly "brain-washed" to restrict the notion of intelligence to the capacities used in solving logical and linguistic problems.

To introduce an alternative point of view, undertake the following "thought experiment." Suspend the usual judgment of what constitutes intelligence and let your thoughts run freely over the capabilities of humans—perhaps those that would be picked out by the proverbial Martian visitor. In this exercise, you are drawn to the brilliant chess player, the world-class violinist, and the champion athlete; such outstanding performers deserve special consideration. Under this experiment, a quite different view of *intelligence* emerges. Are the chess player, violinist, and athlete "intelligent" in these pursuits? If they are, then why do our tests of "intelligence" fail to identify them? If they are not "intelligent," what allows them to achieve such astounding feats? In general, why does the contemporary construct "intelligence" fail to explain large areas of human endeavor?

In this chapter we approach these problems through the theory of multiple intelligences (MI). As the name indicates, we believe that human cognitive competence is better described in terms of a set of abilities, talents, or mental skills, which we call "intelligences." All normal individuals possess each of these skills to some extent; individuals differ in the degree of skill and in the nature of their combination. We believe this theory of intelligence may be more humane and more veridical than alternative views of intelligence and that it more adequately reflects the data of human "intelligent" behavior. Such a theory has important educational implications, including ones for curriculum development.

What Constitutes an Intelligence?

The question of the optimal definition of intelligence looms large in our inquiry. Indeed, it is at the level of this definition that the theory of multiple intelligences diverges from traditional points of view. In a traditional view, intelligence is defined operationally as the ability to answer items on tests of intelligence. The inference from the test scores to some underlying ability is supported by statistical techniques that compare responses of subjects at different ages; the apparent correlation of these test scores across ages and across different tests corroborates the notion that the general faculty of intelligence, *g*, does not change much with age or with training or experience. It is an inborn attribute or faculty of the individual.

Multiple intelligences theory, on the other hand, pluralizes the traditional concept. An intelligence entails the ability to solve problems or fashion prod-

ucts that are of consequence in a particular cultural setting or community. The problem-solving skill allows one to approach a situation in which a goal is to be obtained and to locate the appropriate route to that goal. The creation of a *cultural* product is crucial to such functions as capturing and transmitting knowledge or expressing one's views or feelings. The problems to be solved range from creating an end for a story to anticipating a mating move in chess to repairing a quilt. Products range from scientific theories to musical compositions to successful political campaigns.

MI theory is framed in light of the biological origins of each problem-solving skill. Only those skills that are universal to the human species are treated. Even so, the biological proclivity to participate in a particular form of problem solving must also be coupled with the cultural nurturing of that domain. For example, language, a universal skill, may manifest itself particularly as writing in one culture, as oratory in another culture, and as the secret language of anagrams in a third.

Given the desire of selecting intelligences that are rooted in biology, and that are valued in one or more cultural settings, how does one actually identify an "intelligence"? In coming up with our list, we consulted evidence from several different sources: knowledge about normal development and development in gifted individuals; information about the breakdown of cognitive skills under conditions of brain damage; studies of exceptional populations, including prodigies, idiots savants, and autistic children; data about the evolution of cognition over the millenia; cross-cultural accounts of cognition; psychometric studies, including examinations of correlations among tests; and psychological training studies, particularly measures of transfer and generalization across tasks. Only those candidate intelligences that satisfied all or a majority of the criteria were selected as bona fide intelligences. A more complete discussion of each of these criteria for an "intelligence" and the seven intelligences that have been proposed so far, is found in *Frames of mind* (1983). This book also considers how the theory might be disproven and compares it to competing theories of intelligence.

In addition to satisfying the aforementioned criteria, each intelligence must have an identifiable core operation or set of operations. As a neurally based computational system, each intelligence is activated or "triggered" by certain kinds of internally or externally presented information. For example, one core of musical intelligence is the sensitivity to pitch relations, whereas one core of linguistic intelligence is the sensitivity to phonological features.

An intelligence must also be susceptible to encoding in a symbol system—a culturally contrived system of meaning, which captures and conveys important forms of information. Language, picturing, and mathematics are but three nearly worldwide symbol systems that are necessary for human survival and productivity. The relationship of a candidate intelligence to a human symbol system is no accident. In fact, the existence of a core computational capacity anticipates the existence of a symbol system that exploits that capacity. While it may be possible for an intelligence to proceed without an accompanying symbol system, a primary characteristic of human intelligence may well be its gravitation toward such an embodiment.

The Seven Intelligences

Having sketched the characteristics and criteria of an intelligence, we turn now to a brief consideration of each of the seven intelligences. We begin each sketch with a thumbnail biography of a person who demonstrates an unusual facility with that intelligence. These biographies illustrate some of the abilities that are central to the fluent operation of a given intelligence. Although each biography illustrates a particular intelligence, we do not wish to imply that in adulthood intelligences operate in isolation. Indeed, except for abnormal individuals, intelligences always work in concert, and any sophisticated adult role will involve a melding of several of them. Following each biography we survey the various sources of data that support each candidate as an "intelligence."

Musical Intelligence

> When he was three years old, Yehudi Menuhin was smuggled into the San Francisco Orchestra concerts by his parents. The sound of Louis Persinger's violin so entranced the youngster that he insisted on a violin for his birthday and Louis Persinger as his teacher. He got both. By the time he was ten years old, Menuhin was an international performer (Menuhin, 1977).

Violinist Yehudi Menuhin's musical intelligence manifested itself even before he had touched a violin or received any musical training. His powerful reaction to that particular sound and his rapid progress on the instrument suggest that he was biologically prepared in some way for that endeavor. In this way evidence from child prodigies supports our claim that there is a biological link to a particular intelligence. Other special populations, such as autistic children who can play a musical instrument beautifully but who cannot speak, underscore the independence of musical intelligence.

A brief consideration of the evidence suggests that musical skill passes the other tests for an intelligence. For example, certain parts of the brain play important roles in perception and production of music. These areas are characteristically located in the right hemisphere, although musical skill is not as clearly "localized," or located in a specifiable area, as language. Although the particular susceptibility of musical ability to brain damage depends on the degree of training and other individual differences, there is clear evidence for "amusia" or loss of musical ability.

Music apparently played an important unifying role in Stone Age (Paleolithic) societies. Birdsong provides a link to other species. Evidence from various cultures supports the notion that music is a universal faculty. Studies of infant development suggest that there is a "raw" computational ability in early childhood. Finally, musical notation provides an accessible and lucid symbol system.

In short, evidence to support the interpretation of musical ability as an "intelligence" comes from many different sources. Even though musical skill is not typically considered an intellectual skill like mathematics, it qualifies under our criteria. By definition it deserves consideration; and in view of the data, its inclusion is empirically justified.

Bodily-Kinesthetic Intelligence

> Fifteen-year-old Babe Ruth played third base. During one game his team's pitcher was doing very poorly and Babe loudly criticized him from third base. Brother Mathias, the coach, called out, "Ruth, if you know so much about it, YOU pitch!" Babe was surprised and embarrassed because he had never pitched before, but Brother Mathias insisted. Ruth said later that at the very moment he took the pitcher's mound, he KNEW he was supposed to be a pitcher and that it was "natural" for him to strike people out. Indeed, he went on to become a great major league pitcher (and, of course, attained legendary status as a hitter) (Connor, 1982).

Like Menuhin, Babe Ruth was a child prodigy who recognized his "instrument" immediately upon his first exposure to it. This recognition occurred in advance of formal training.

Control of bodily movement is, of course, localized in the motor cortex, with each hemisphere dominant or controlling bodily movements on the contralateral side. In right-handers, the dominance for such movement is ordinarily found in the left hemisphere. The ability to perform movements when directed to do so can be impaired even in individuals who can perform the same movements reflexively or on a nonvoluntary basis. The existence of specific *apraxia* constitutes one line of evidence for a bodily-kinesthetic intelligence.

The evolution of specialized body movements is of obvious advantage to the species, and in humans this adaptation is extended through the use of tools. Body movement undergoes a clearly defined developmental schedule in children. And there is little question of its universality across cultures. Thus it appears that bodily-kinesthetic "knowledge" satisfies many of the criteria for an intelligence.

The consideration of bodily-kinesthetic knowledge as "problem solving" may be less intuitive. Certainly carrying out a mime sequence or hitting a tennis ball is not solving a mathematical equation. And yet, the ability to use one's body to express an emotion (as in a dance), to play a game (as in a sport), or to create a new product (as in devising an invention) is evidence of the cognitive features of body usage. The specific computations required to solve a particular bodily-kinesthetic *problem*, hitting a tennis ball, are summarized by Tim Gallwey:

> At the moment the ball leaves the server's racket, the brain calculates approximately where it will land and where the racket will intercept it. This calculation includes the initial velocity of the ball, combined with an input for the progressive decrease in velocity and the effect of wind and after the bounce of the ball. Simultaneously, muscle orders are given: not just once, but constantly with refined and updated information. The muscles must cooperate. A movement of the feet occurs, the racket is taken back, the face of the racket kept at a constant angle. Contact is made at a precise point that depends on whether the order was given to hit down the line or cross-court, an order not given until after a split-second analysis of the movement and balance of the opponent.
>
> To return an average serve, you have about one second to do this. To hit the ball at all is remarkable and yet not uncommon. The truth is that

> everyone who inhabits a human body possesses a remarkable creation (Gallwey, 1976).

Logical-Mathematical Intelligence

In 1983 Barbara McClintock won the Nobel Prize in medicine or physiology for her work in microbiology. Her intellectual powers of deduction and observation illustrate one form of logical-mathematical intelligence that is often labeled "scientific thinking." One incident is particularly illuminating. While a researcher at Cornell in the 1920s McClintock was faced one day with a problem: while *theory* predicted 50 percent pollen sterility in corn, her research assistant (in the "field") was finding plants that were only 25 to 30 percent sterile. Disturbed by this discrepancy, McClintock left the cornfield and returned to her office where she sat for half an hour, thinking:

> Suddenly I jumped up and ran back to the (corn) field. At the top of the field (the others were still at the bottom) I shouted "Eureka, I have it! I know what the 30% sterility is!" . . . They asked me to prove it. I sat down with a paper bag and a pencil and I started from scratch, which I had not done at all in my laboratory. It had all been done so fast; the answer came and I ran. Now I worked it out step by step—it was an intricate series of steps—and I came out with [the same result]. [They] looked at the material and it was exactly as I'd said it was; it worked out exactly as I had diagrammed it. Now, why did I know, without having done it on paper? Why was I so sure? (Keller, 1983, p. 104).

This anecdote illustrates two essential facts of the logical-mathematical intelligence. First, in the gifted individual, the process of problem solving is often remarkably rapid—the successful scientist copes with many variables at once and creates numerous hypotheses that are each evaluated and then accepted or rejected in turn.

The anecdote also underscores the *nonverbal* nature of the intelligence. A solution to a problem can be constructed *before* it is articulated. In fact, the solution process may be totally invisible, even to the problem solver. This need not imply, however, that discoveries of this sort—the familiar "Aha!" phenomenon—are mysterious, intuitive, or unpredictable. The fact that it happens more frequently to some people (perhaps Nobel Prize winners) suggests the opposite. We interpret this as the work of the logical-mathematical intelligence.

Along with the companion skill of language, logical-mathematical reasoning provides the principal basis for IQ tests. This form of intelligence has been heavily investigated by traditional psychologists, and it is the archetype of "raw intelligence" or the problem-solving faculty that purportedly cuts across domains. It is perhaps ironic, then, that the actual mechanism by which one arrives at a solution to a logical-mathematical problem is not as yet properly understood.

This intelligence is supported by our empirical criteria as well. Certain areas of the brain are more prominent in mathematical calculation than others. There are idiots savants who perform great feats of calculation even though they re-

main tragically deficient in most other areas. Child prodigies in mathematics abound. The development of this intelligence in children has been carefully documented by Jean Piaget and other psychologists.

Linguistic Intelligence

> At the age of ten, T. S. Eliot created a magazine called "Fireside" to which he was the sole contributor. In a three-day period during his winter vacation, he created eight complete issues. Each one included poems, adventure stories, a gossip column, and humor. Some of this material survives and it displays the talent of the poet (see Soldo, 1982).

As with the logical intelligence, calling linguistic skill an "intelligence" is consistent with the stance of traditional psychology. Linguistic intelligence also passes our empirical tests. For instance, a specific area of the brain, called "Broca's Area," is responsible for the production of grammatical sentences. A person with damage to this area can understand words and sentences quite well but has difficulty putting words together in anything other than the simplest of sentences. At the same time, other thought processes may be entirely unaffected.

The gift of language is universal, and its development in children is strikingly constant across cultures. Even in deaf populations where a manual sign language is not explicitly taught, children will often "invent" their own manual language and use it surreptitiously! We thus see how an intelligence may operate independently of a specific input modality or output channel.

Spatial Intelligence

> Navigation around the Caroline Islands in the South Seas is accomplished without instruments. The position of the stars, as viewed from various islands, the weather patterns, and water color are the only sign posts. Each journey is broken into a series of segments; and the navigator learns the position of the stars within each of these segments. During the actual trip the navigator must envision mentally a reference island as it passes under a particular star and from that he computes the number of segments completed, the proportion of the trip remaining, and any corrections in heading that are required. The navigator cannot *see* the islands as he sails along; instead he maps their locations in his mental "picture" of the journey (Gardner, 1983).

Spatial problem solving is required for navigation and in the use of the notational system of maps. Other kinds of spatial problem solving are brought to bear in visualizing an object seen from a different angle and in playing chess. The visual arts also employ this intelligence in the use of space.

Evidence from brain research is clear and persuasive. Just as the left hemisphere has, over the course of evolution, been selected as the site of linguistic processing in right-handed persons, the right hemisphere proves to be the site most crucial for spatial processing. Damage to the right posterior regions causes impairment of the ability to find one's way around a site, to recognize faces or scenes, or to notice fine details.

Patients with damage specific to regions of the right hemisphere will attempt to compensate for their spacial deficits with linguistic strategies. They will try to reason aloud, to challenge the task, or even make up answers. But such nonspatial strategies are rarely successful.

Blind populations provide an illustration of the distinction between the spatial intelligence and visual perception. A blind person can recognize shapes by an indirect method: running a hand along the object translates into length of time of movement, which in turn is translated into the size of the object. For the blind person, the perceptual system of the tactile modality parallels the visual modality in the seeing person. The analogy between the spatial reasoning of the blind and the linguistic reasoning of the deaf is notable.

There are few child prodigies among visual artists, but there are idiots savants such as Nadia (Selfe, 1977). Despite a condition of severe autism, this preschool child made drawings of the most remarkable representational accuracy and finesse.

Interpersonal Intelligence

With little formal training in special education and nearly blind herself, Anne Sullivan began the intimidating task of instructing a blind and deaf seven-year-old Helen Keller. Sullivan's efforts at communication were complicated by the child's emotional struggle with the world around her. At their first meal together, this scene occurred:

> Annie did not allow Helen to put her hand into Annie's plate and take what she wanted, as she had been accustomed to do with her family. It became a test of wills—hand thrust into plate, hand firmly put aside. The family, much upset, left the dining room. Annie locked the door and proceeded to eat her breakfast while Helen lay on the floor kicking and screaming, pushing and pulling at Annie's chair. [After half an hour] Helen went around the table looking for her family. She discovered no one else was there and that bewildered her. Finally, she sat down and began to eat her breakfast, but with her hands. Annie gave her a spoon. Down on the floor it clattered, and the contest of wills began anew (Lash, 1980, p. 52).

Anne Sullivan sensitively responded to the child's behavior. She wrote home: "The greatest problem I shall have to solve is how to discipline and control her without breaking her spirit. I shall go rather slowly at first and try to win her love."

In fact, the first "miracle" occurred two weeks later, well before the famous incident at the pumphouse. Annie had taken Helen to a small cottage near the family's house, where they could live alone. After seven days together, Helen's personality suddenly underwent a profound change—the therapy had worked:

> My heart is singing with joy this morning. A miracle has happened! The wild little creature of two weeks ago has been transformed into a gentle child (p. 54).

It was just two weeks after this that the first breakthrough in Helen's grasp of language occurred; and from that point on, she progressed with incredible

speed. The key to the miracle of language was Anne Sullivan's insight into the *person* of Helen Keller.

Interpersonal intelligence builds on a core capacity to notice distinctions among others; in particular, contrasts in their moods, temperaments, motivations, and intentions. In more advanced forms, this intelligence permits a skilled adult to read the intentions and desires of others, even when these have been hidden. This skill appears in a highly sophisticated form in religious or political leaders, teachers, therapists, and parents. The Helen Keller–Anne Sullivan story suggests that this interpersonal intelligence does not depend on language.

All indices in brain research suggest that the frontal lobes play a prominent role in interpersonal knowledge. Damage in this area can cause profound personality changes while leaving other forms of problem solving unharmed—a person is often "not the same person" after such an injury.

Alzheimer's disease, a form of presenile dementia, appears to attack posterior brain zones with a special ferocity, leaving spatial, logical, and linguistic computations severly impaired. Yet, Alzheimer's patients will often remain well groomed, socially proper, and continually apologetic for their errors. In contrast, Pick's disease, another variety of presenile dementia that is more frontally oriented, entails a rapid loss of social graces.

Biological evidence for interpersonal intelligence encompasses two additional factors often cited as unique to humans. One factor is the prolonged childhood of primates, including the close attachment to the mother. In those cases where the mother is removed from early development, normal interpersonal development is in serious jeopardy. The second factor is the relative importance in humans of social interaction. Skills such as hunting, tracking, and killing in prehistoric societies required participation and cooperation of large numbers of people. The need for group cohesion, leadership, organization, and solidarity follows naturally from this.

Intrapersonal Intelligence

In an essay called "A Sketch of the Past," written almost as a diary entry, Virginia Woolf discusses the "cotton wool of existence"—the various mundane events of life. She contrasts this "cotton wool" with three specific and poignant memories from her childhood: a fight with her brother, seeing a particular flower in the garden, and hearing of the suicide of a past visitor:

> These are three instances of exceptional moments. I often tell them over, or rather they come to the surface unexpectedly. But now for the first time I have written them down, and I realize something that I have never realized before. Two of these moments ended in a state of despair. The other ended, on the contrary, in a state of satisfaction.
>
> The sense of horror (in hearing of the suicide) held me powerless. But in the case of the flower, I found a reason; and was thus able to deal with the sensation. I was not powerless.
>
> Though I still have the peculiarity that I receive these sudden shocks, they are now always welcome; after the first surprise, I always feel instantly that they are particularly valuable. And so I go on to suppose

> that the shock-receiving capacity is what makes me a writer. I hazard the explanation that a shock is at once in my case followed by the desire to explain it. I feel that I have had a blow; but it is not, as I thought as a child, simply a blow from an enemy hidden behind the cotton wool of daily life; it is or will become a revelation of some order; it is a token of some real thing behind appearances; and I make it real by putting it into words (Woolf, 1976, pp. 69–70).

This quotation vividly illustrates the intrapersonal intelligence—knowledge of the internal aspects of a person: access to one's own feeling life, one's range of emotions, the capacity to effect discriminations among these emotions and eventually to label them and to draw upon them as a means of understanding and guiding one's own behavior. A person with good intrapersonal intelligence has a viable and effective model of himself or herself. Since this intelligence is the most private, it requires evidence from language, music, or some other more expressive form of intelligence if the observer is to detect it at work. In the above quotation, for example, linguistic intelligence is drawn upon to convey intrapersonal knowledge; it embodies the interaction of intelligences, a common phenomenon to which we will return later.

We see the familiar criteria at work in the intrapersonal intelligence. As with the interpersonal intelligence, the frontal lobes play a central role in personality change. Injury to the lower area of the frontal lobes is likely to produce irritability or euphoria; while injury to the higher regions is more likely to produce indifference, listlessness, slowness, and apathy—a kind of depressive personality. In such "frontal-lobe" individuals, the other cognitive functions often remain preserved. In contrast, among aphasics who have recovered sufficiently to describe their experiences, we find consistent testimony: while there may have been a diminution of general alertness and considerable depression about the condition, the individual in no way felt himself to be a different person. He recognized his own needs, wants, and desires and tried as best he could to achieve them.

The autistic child is a prototypical example of an individual with impaired intrapersonal intelligence; indeed, the child may not even be able to refer to himself. At the same time, such children often exhibit remarkable abilities in the musical, computational, spatial, or mechanical realms.

Evolutionary evidence for an intrapersonal faculty is more difficult to come by, but we might speculate that the capacity to transcend the satisfaction of instinctual drives is relevant. This becomes increasingly important in a species not perennially involved in the struggle for survival.

In sum, then, both interpersonal and intrapersonal faculties pass the tests of an intelligence. They both feature problem-solving endeavors with significance for the individual and the species. Interpersonal intelligence allows one to understand and work with others; intrapersonal intelligence allows one to understand and work with oneself. In the individual's sense of self, one encounters a melding of inter- and intrapersonal components. Indeed, the sense of self emerges as one of the most marvelous of human inventions—a symbol that represents all kinds of information about a person and that is at the same time an invention that all individuals construct for themselves.

Summary: The Unique Contributions of the Theory

As human beings, we all have a repertoire of skills for solving different kinds of problems. Our investigation has begun, therefore, with a consideration of these problems, the contexts they are found in, and the culturally significant products that are the outcome. We have not approached "intelligence" as a reified human faculty that is brought to bear in literally any problem setting; rather, we have begun with the problems that humans *solve* and worked back to the "intelligences" that must be responsible.

Evidence from brain research, human development, evolution, and cross-cultural comparisons was brought to bear in our search for the relevant human intelligences: a candidate was included only if reasonable evidence to support its membership was found across these diverse fields. Again, this tack differs from the traditional one: since no candidate faculty is *necessarily* an intelligence, we could choose on a motivated basis. In the traditional approach to "intelligence," there is no opportunity for this type of empirical decision.

We have also determined that these multiple human faculties, the intelligences, are to a significant extent *independent*. For example, research with brain-damaged adults repeatedly demonstrates that particular faculties can be lost while others are spared. This independence of intelligences implies that a particularly high level of ability in one intelligence, say mathematics, does not require a similarly high level in another intelligence, like language or music. This independence of intelligences contrasts sharply with traditional measures of IQ that find high correlations among test scores. We speculate that the usual correlations among subtests of IQ tests come about because all of these tasks in fact measure the ability to respond rapidly to items of a logical-mathematical or linguistic sort; we believe that these correlations would be substantially reduced if one were to survey in a contextually appropriate way the full range of human problem-solving skills.

Until now, we have supported the fiction that adult roles depend largely on the flowering of a single intelligence. In fact, however, nearly every cultural role of any degree of sophistication requires a combination of intelligences. Thus, even an apparently straightforward role, like playing the violin, transcends a reliance on simple musical intelligence. To become a successful violinist requires bodily-kinesthetic dexterity and the interpersonal skills of relating to an audience and, in a different way, choosing a manager; quite possibly it involves an intrapersonal intelligence as well. Dance requires skills in bodily-kinesthetic, musical, interpersonal, and spatial intelligences in varying degrees. Politics requires an interpersonal skill, a linguistic facility, and perhaps some logical aptitude. Inasmuch as nearly every cultural role requires several intelligences, it becomes important to consider individuals as a collection of aptitudes rather than as having a singular problem-solving faculty that can be measured directly through pencil-and-paper tests. Even given a relatively small number of such intelligences, the diversity of human ability is created through the differences in these profiles. In fact, it may well be that the "total is greater than the sum of the parts." An individual may not be particularly gifted in any intelligence; and yet, because of a particular combination or blend of skills, he or she may be able to fill some niche uniquely well. Thus it is of paramount

importance to assess the particular combination of skills that may earmark an individual for a certain vocational or avocational niche.

Implications for Education

The theory of multiple intelligences was developed as an account of human cognition that can be subjected to empirical tests. In addition, the theory seems to harbor a number of educational implications that are worth consideration. In the following discussion we will begin by outlining what appears to be the natural developmental trajectory of an intelligence. Turning then to aspects of education, we will comment on the role of nurturing and explicit instruction in this development. From this analysis we find that assessment of intelligences can play a crucial role in curriculum development.

The Natural Growth of an Intelligence: A Developmental Trajectory

Since all intelligences are part of the human genetic heritage, at some basic level each intelligence is manifested universally, independent of education and cultural support. Exceptional populations aside for the moment, *all* humans possess certain core abilities in each of the intelligences.

The natural trajectory of development in each intelligence begins with *raw patterning ability*, for example, the ability to make tonal differentiations in musical intelligence or to appreciate three-dimensional arrangements in spatial intelligence. These abilities appear universally; they may also appear at a heightened level in that part of the population that is "at promise" in that domain. The "raw" intelligence predominates during the first year of life.

Intelligences are glimpsed through different lenses at subsequent points in development. In the subsequent stage, the intelligence is encountered through a *symbol system*: language is encountered through sentences and stories, music through songs, spatial understanding through drawings, bodily-kinesthetic through gesture or dance, and so on. At this point children demonstrate their abilities in the various intelligences through their grasp of various symbol systems. Yehudi Menuhin's response to the sound of the violin illustrates the musical intelligence of a gifted individual coming in contact with a particular aspect of the symbol system.

As development progresses, each intelligence together with its accompanying symbol system is represented in a *notational system*. Mathematics, mapping, reading, music notation, and so on, are second-order symbol systems in which the marks on paper come to stand for symbols. In our culture, these notational systems are typically mastered in a formal educational setting.

Finally, during adolescence and adulthood, the intelligences are expressed through the range of *vocational and avocational pursuits*. For example, the logical-mathematical intelligence, which began as sheer pattern ability in infancy and developed through symbolic mastery of early childhood and the notations of the school years, achieves mature expression in such roles as mathematician, accountant, scientist, cashier. Similarly, the spatial intelligence passes from the mental maps of the infant, to the symbolic operations required in drawings and

the notational systems of maps, to the adult roles of navigator, chess player, and topologist.

Although all humans partake of each intelligence to some degree, certain individuals are said to be "at promise." They are highly endowed with the core abilities and skills of that intelligence. This fact becomes important for the culture as a whole, since, in general, these exceptionally gifted individuals will make notable advances in the cultural manifestations of that intelligence. It is not important that *all* members of the Puluwat tribe demonstrate precocious spatial abilities needed for navigation by the stars, nor is it necessary for all Westerners to master mathematics to the degree necessary to make a significant contribution to theoretical physics. So long as the individuals "at promise" in particular domains are located efficiently, the overall knowledge of the group will be advanced in all domains.

While some individuals are "at promise" in an intelligence, others are "at risk." In the absence of special aids, those at risk in an intelligence will be most likely to fail tasks involving that intelligence. Conversely, those at promise will be most likely to succeed. It may be that intensive intervention at an early age can bring a larger number of children to an "at promise" level.

The special developmental trajectory of an individual at promise varies with intelligence. Thus, mathematics and music are characterized by the early appearance of gifted children who perform relatively early at or near an adult level. In contrast, the personal intelligences appear to arise much more gradually; prodigies are rare. Moreover, mature performance in one area does not imply mature performance in another area, just as gifted achievement in one does not imply gifted achievement in another.

Implications of the Developmental Trajectory for Education

Because the intelligences are manifested in different ways at different developmental levels, both assessment and nurturing need to occur in apposite ways. What nurtures in infancy would be inappropriate at later stages, and vice versa. In the preschool and early elementary years, instruction should emphasize opportunity. It is during these years that children can discover something of their own peculiar interests and abilities.

In the case of very talented children, such discoveries often happen by themselves through spontaneous "crystallizing experiences" (Walters & Gardner, 1986). When such experiences occur, often in early childhood, an individual reacts overtly to some attractive quality or feature of a domain. Immediately the individual undergoes a strong affective reaction; he or she feels a special affinity to that domain, as did Menuhin when he first heard the violin at an orchestral concert. Thereafter, in many cases, the individual persists working in the domain, and, by drawing on a powerful set of appropriate intelligences, goes on to achieve high skill in that domain in relatively quick compass.

In the case of the most powerful talents, such crystallizing experiences seem difficult to prevent; and they may be especially likely to emerge in the domains of music and mathematics. However, specifically designed encounters with materials, equipment, or other people can help a youngster discover his or her own métier.

During the school-age years, some mastery of notational systems is essential in our society. The self-discovery environment of early schooling cannot provide the structure needed for the mastery of specific notational systems like the sonata form or algebra. In fact, during this period some tutelage is needed by virtually all children. One problem is to find the right form, since group tutelage can be helpful in some instances and harmful in others. Another problem is to orchestrate the connection between practical knowledge and the knowledge embodied in symbolic systems and notational systems.

Finally, in adolescence, most students must be assisted in their choice of careers. This task is made more complex by the manner in which intelligences interact in many cultural roles. For instance, being a doctor certainly requires logical-mathematical intelligence; but while the general practitioner should have strong interpersonal skills, the surgeon needs bodily-kinesthetic dexterity. Internships, apprenticeships, and involvement with the actual materials of the cultural role become critical at this point in development.

Several implications for explicit instruction can be drawn from this analysis. First, the role of instruction in relation to the manifestation of an intelligence changes across the developmental trajectory. The enriched environment appropriate for the younger years is less crucial for adolescents. Conversely, explicit instruction in the notational system, appropriate for older children, is largely inappropriate for younger ones.

Explicit instruction must be evaluated in light of the developmental trajectories of the intelligences. Students benefit from explicit instruction only if the information or training fits into their specific place on the developmental progression. A particular kind of instruction can be either too early at one point or too late at another. For example, Suzuki training in music pays little attention to the notational system, while providing a great deal of support or scaffolding for learning the fine points of instrumental technique. While this emphasis may be very powerful for training preschool children, it can produce stunted musical development when imposed at a late point on the developmental trajectory. Such a highly structured instructional environment can accelerate progress and produce a larger number of children "at promise," but in the end it may ultimately limit choices and inhibit self-expression.

An exclusive focus on linguistic and logical skills in formal schooling can shortchange individuals with skills in other intelligences. It is evident from inspection of adult roles, even in language-dominated Western society, that spatial, interpersonal, or bodily-kinesthetic skills often play key roles. Yet linguistic and logical skills form the core of most diagnostic tests of "intelligence" and are placed on a pedagogical pedestal in our schools.

The Large Need: Assessment

The general pedagogical program described here presupposes accurate understanding of the profile of intelligences of the individual learner. Such a careful assessment procedure allows informed choices about careers and avocations. It also permits a more enlightened search for remedies for difficulties. Assessment of deficiencies can predict difficulties the learner will have; moreover, it can suggest alternative routes to an educational goal (learning mathematics via spatial relations; learning music through linguistic techniques).

Assessment, then, becomes a central feature of an educational system. We believe that it is essential to depart from standardized testing. We also believe that standard pencil-and-paper short-answer tests sample only a small proportion of intellectual abilities and often reward a certain kind of decontextualized facility. The means of assessment we favor should ultimately search for genuine problem-solving or product-fashioning skills in individuals across a range of materials.

An assessment of a particular intelligence (or set of intelligences) should highlight problems that can be solved *in the materials of that intelligence*. That is, mathematical assessment should present problems in mathematical settings. For younger children, these could consist of Piagetian-style problems in which talk is kept to a minimum. For older children, derivation of proofs in a novel numerical system might suffice. In music, on the other hand, the problems would be embedded in a musical system. Younger children could be asked to assemble tunes from individual musical segments. Older children could be shown how to compose a rondo or fugue from simple motifs.

An important aspect of assessing intelligences must include the individual's ability to solve problems or create products using the materials of the intellectual medium. Equally important, however, is the determination of which intelligence is favored when an individual has a choice. One technique for getting at this proclivity is to expose the individual to a sufficiently complex situation that can stimulate several intelligences; or to provide a set of materials drawn from different intelligences and determine toward which one an individual gravitates and how deeply he or she explores it.

As an example, consider what happens when a child sees a complex film in which several intelligences figure prominently: music, people interacting, a maze to be solved, or a particular bodily skill, may all compete for attention. Subsequent "debriefing" with the child should reveal the features to which the child paid attention; these will be related to the profile of intelligences in that child. Or consider a situation in which children are taken into a room with several different kinds of equipment and games. Simple measures of the regions in which children spend time and the kinds of activities they engage in should yield insights into the individual child's profile of intelligence.

Tests of this sort differ in two important ways from the traditional measures of "intelligence." First, they rely on materials, equipment, interviews, and so on to generate the problems to be solved; this contrasts with the traditional pencil-and-paper measures used in intelligence testing. Second, results are reported as part of an individual profile of intellectual propensities, rather than as a single index of intelligence or rank within the population. In contrasting strengths and weaknesses, they can suggest options for future learning.

Scores are not enough. This assessment procedure should suggest to parents, teachers, and, eventually, to children themselves, the sorts of activities that are available at home, in school, or in the wider community. Drawing on this information, children can bolster their own particular sets of intellectual weaknesses or combine their intellectual strengths in a way that is satisfying vocationally and avocationally.

Coping with the Plurality of Intelligences

Under the multiple intelligences theory, an intelligence can serve both as the *content* of instruction and the *means* or medium for communicating that content. This state of affairs has important ramifications for instruction. For example, suppose that a child is learning some mathematical principle but is not skilled in logical-mathematical intelligence. That child will probably experience some difficulty during the learning process. The reason for the difficulty is straightforward: the mathematical principle to be learned (the content) exists only in the logical-mathematical world and it ought to be communicated through mathematics (the medium). That is, the mathematical principle cannot be translated *entirely* into words (a linguistic medium) or spatial models (a spatial medium). At some point in the learning process, the mathematics of the principle must "speak for itself." In our present case, it is at just this level that the learner experiences difficulty—the learner (who is not especially "mathematical") and the problem (which is very much "mathematical") are not in accord. Mathematics, as a *medium*, has failed.

Although this situation is a necessary conundrum in light of multiple intelligences theory, we can propose various solutions. In the present example, the teacher must attempt to find an alternative route to the mathematical content—a metaphor in another medium. Language is perhaps the most obvious alternative, but spatial modeling and even a bodily-kinesthetic metaphor may prove appropriate in some cases. In this way, the student is given a *secondary* route to the solution to the problem, perhaps through the medium of an intelligence that is relatively strong for that individual.

Two features of this hypothetical scenario must be stressed. First, in such cases, the secondary route—the language, spatial model, or whatever—is at best a metaphor or translation. It is not mathematics itself. And at some point, the learner must translate back into the domain of mathematics. Without this translation, what is learned tends to remain at a relatively superficial level; cookbook-style mathematical performance results from following instructions (linguistic translation) without understanding why (mathematics retranslation).

Second, the alternative route is not guaranteed. There is no *necessary* reason why a problem in one domain *must be translatable* into a metaphorical problem in another domain. Successful teachers find these translations with relative frequency; but as learning becomes more complex, the likelihood of a successful translation may diminish.

While multiple intelligences theory is consistent with much empirical evidence, it has not been subjected to strong experimental tests within psychology. Within the area of education, the applications of the theory are currently being examined in many projects. Our hunches will have to be revised many times in light of actual classroom experience. Still there are important reasons for considering the theory of multiple intelligences and its implications for education. First of all, it is clear that many talents, if not intelligences, are overlooked nowadays; individuals with these talents are the chief casualties of the single-minded, single-funneled approach to the mind. There are many unfilled or poorly filled niches in our society and it would be opportune to guide individuals with the right set of abilities to these billets. Finally, our world is beset with problems; to have any chance of solving them, we must make the very

best use of the intelligences we possess. Perhaps recognizing the plurality of intelligences and the manifold ways in which human individuals may exhibit them is an important first step.

Acknowledgments

The research reported in this chapter was supported by grants from the Bernard Van Leer Foundation of The Hague, the Spencer Foundation of Chicago, and the Carnegie Corporation of New York. We are grateful to Mara Krechevsky, who gave many helpful comments on earlier drafts.

References

Connor, A. (1982). *Voices from Cooperstown*. New York: Collier. (Based on a quotation taken from *The Babe Ruth story*, Babe Ruth & Bob Considine. New York: Dutton, 1948.)

Gallwey, T. (1976). *Inner tennis*. New York: Random House.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Jencks, C. (1972). *Inequality*. New York: Basic Books.

Keller, E. (1983). *A feeling for the organism*. Salt Lake City: W. H. Freeman.

Lash, J. (1980). *Helen and teacher: The story of Helen Keller and Anne Sullivan Macy*. New York: Delacorte.

Menuhin, Y. (1977). *Unfinished journey*. New York: Knopf.

Selfe, L. (1977). *Nadia: A case of extraordinary drawing ability in an autistic child*. New York: Academic Press.

Soldo, J. (1982). Jovial juvenilia: T. S. Eliot's first magazine. *Biography, 5*, 25–37.

Walters, J., & Gardner, H. (1986). The crystallizing experience: Discovering an intellectual gift. In R. Sternberg & J. Davidson (Eds.), Conceptions of giftedness (pp. 306–31). New York: Cambridge University Press.

Woolf, V. (1976). *Moments of being*. Sussex: The University Press.

Chapter 36

Individual Differences in Cognition

R. Kim Guenther

People differ with respect to their intellectual capabilities. Historically, the attempt to measure differences in intellectual ability has been the most conspicuous and influential branch of cognitive psychology.

Perspectives on Individual Differences in Intelligence: Hereditarian, Unitary Models versus Multifaceted, Domain-Specific Models of Intelligence

The assumptions historically made by researchers in the intelligence testing movement constitute a theory of intelligence that Steven Jay Gould calls the *hereditarian theory* (Gould, 1981; also Mackintosh, 1986). The hereditarian theory of intelligence makes two separate claims. First, it claims that intelligence is *unitary*—it is a reflection of an all-purpose system or process that permeates all intellectual activity. Another way of making this claim is to say that intelligence is *generic*. An implication of the generic notion is that intelligence is measurable using tests that are meaningfully converted into numbers that reflect the amount of intelligence a person possesses. The second claim, from which the hereditarian theory derives its name, is that the primary basis of intellectual differences among people is to be found in the genes they inherit; that is, intelligence is primarily *genetically determined*. Although these claims are logically distinct (intelligence could be unitary but differences among people could still be due primarily to environmental differences), historically they have been associated.

The main theme of this chapter will be a comparison between the unitary or generic view of individual cognitive differences, on the one hand, and a *domain-specific* or *multifaceted* view of individual cognitive differences, on the other (Gardner, 1983). The multifaceted view claims that people may display superior talent or skill in one intellectual domain without necessarily being superior in other domains. As I did in chapter 8: Problem Solving, I will champion here the domain-specific approach to individual intellectual differences. I will also discuss the evidence for a genetic basis for intellectual differences and try to make clear what are and are not reasonable implications of this evidence. Included in the section on the genetic basis of intelligence is a discussion of sex differences in cognitive skills.

36.1 Historical Background and the Rise of the Heredtarian Theory of Intelligence

A confluence of several developments taking place in the 1800s led to an interest in the measurement of individual differences in cognition, culminating in

From chapter 9 in *Human Cognition* (New York: Prentice-Hall, 1998), 313–346. Reprinted with permission.

the creation of *intelligence quotient (IQ)* tests around the turn of the 20th century. One development was the theory of evolution, which focuses on individual differences. For traits like abstract reasoning or language to evolve in a species, members of predecessor species must differ from one another on that trait. Only then can natural selection produce an increase in the number of individuals possessing the more adaptive trait. A second development was the growing acceptance of materialism—the view that what we label mental activity reflects only brain processes. In this view, any intellectual differences between people must also be reflected in differences in their brains. A third development was the rise of psychological experimentation and measurement. Sophisticated techniques for investigating and quantifying human behavior were being developed in the experimental laboratories of Europe and North America. Finally, the industrialized nations had become committed to universal education. But not everyone seemed to profit very much by formal education. Consequently, educators became interested in identifying students who might need special educational intervention.

The Rise of the Intelligence Testing Movement

Francis Galton Francis Galton, Darwin's cousin and one of the founders of the intelligence testing movement, was a bright, independently wealthy man who had a passion for measuring things. He was the first to suggest that fingerprints be used for personal identification. He measured the degree of boredom at scientific lectures, and tried to find out which country had the most beautiful women.

Galton, along with his friend Karl Pearson (1867–1936), devised the concept and formula for *correlation* (see Boring, 1950; Gould, 1981; Hergenhahn, 1986). As it turns out, the concept of correlation is extremely important to the research on intelligence. Correlation is a measure of the degree to which two measurements are linearly related. Correlations range between +1 and −1. A positive correlation indicates that when scores on one measure increase, scores on the other measure tend to increase as well. A negative correlation indicates that when scores on one measure increase, scores on the other measure tend to decrease. A lack of correlation between two measures means that when scores on one measure increase, scores on the other measure tend neither to increase nor decrease. Height and weight are positively correlated—people who are tall also tend to be people who weigh more. Smoking and longevity are negatively correlated—people who smoke more tend to live fewer years. The last digit of one's social security number and one's annual income in dollars are not correlated—people with higher last digits are not likely to earn more money or less money.

It is important to note that just because two measures are correlated does not mean that there is a causal relationship between them. However, if there is a causal relationship, it is certain that the two measures will be correlated. There is a positive correlation between the speed with which a sprinter runs and the number of wins in a track meet. Here the faster speed is the cause of the winning. But there is also a positive correlation between the number of ice cream cones consumed in New York City on any given day and the number of deaths

in Bombay, India on the same given day. Obviously, though, the eating of ice cream cones in New York does not cause people in Bombay to die; rather, both measures probably reflect global climate. When it is hot in the Northern Hemisphere, people in New York eat ice cream cones and people in Bombay endure heat and disease. Many correlations are simply coincidental. The gross national product of the United States in any given year is positively correlated to the distance between the North American continent and the European continent—both are increasing over time.

Based on his correlational and measuring techniques, Galton (1883) decided that intelligence is primarily a reflection of energy and the perceptual acuteness of the senses. Intelligent people, thought Galton, were especially good at perceptually discriminating between similar stimuli, such as between two similar colors differing only slightly in frequency. In 1884 he set up an anthropometric laboratory at the International Exposition where visitors, by paying a threepence, could have their skulls measured and have various tests taken of their perceptual functions. Some of the tests included judging the relative weight of a series of identical-looking objects, trying to detect very high frequency sounds, and reacting as quickly as possible to an auditory stimulus by punching a bag. This laboratory, later transferred to South Kensington Museum in London, constituted the first large-scale testing of individual differences.

Galton claimed that mentally retarded people did not discriminate heat, cold, and pain as well as "normal" people, and used this finding to bolster his argument that sensory discriminatory capacity underlies intelligence (Galton, 1883). Other research seemed to show that children classified by their teachers as "bright" tended to have faster reaction times than children classified as below average (Gilbert, 1894). Galton's procedures for measuring intelligence were adopted by James Cattell (1860–1944), who administered them to college students in the United States (Cattell, 1890).

Later research discredited some of Galton's ideas, when it was shown that an individual's performance on sensory and reaction time tests showed little relationship from test to test, and was unrelated to grades in school or to a teacher's estimates of intelligence (e.g., Wissler, 1901). More recent research (discussed below) suggests that there might be a modest relationship between performance on sensory or reaction-time tests and other measures of intellectual prowess.

Galton's interest in evolution led him to study the possibility that intelligence runs in families. Based on a study of families of people who were highly acclaimed scientists, artists, writers, and politicians, Galton found that children of illustrious people were more likely to be illustrious than children of ordinary folks (Galton, 1884). Galton concluded that the basis of high intelligence was favorable genes that the illustrious passed on to their offspring. Galton advocated a form of eugenics, in which the government would pay highly intelligent people to marry and bear children.

Alfred Binet Alfred Binet (1857–1911), one of the founders of experimental psychology in France, conducted research on hypnotism, cognitive development, memory, and creativity. Some of his work with children was similar to that later conducted by Jean Piaget (see Boring, 1950; Gould, 1981; Hergenhahn, 1986).

In 1903 Binet and Theodore Simon (1878–1961) were commissioned by the French government to develop a test that could identify learning disabled or mentally retarded children, so that they could be given special education. At the time, tests based on Galton's theories were used, but, as discussed before, some research seemed to discredit Galton's ideas about the basis of individual differences. Besides, as Binet noted, children with vision and hearing impairments would be erroneously classified as retarded. Binet proposed instead that more complex tests of reasoning, motor performance, spatial thinking, and memory be used to assess a child's cognitive abilities. Binet and Simon's tests included reasoning problems, reflecting Binet's belief that the intelligent person was one who showed reasoned judgments when confronted with problems (Binet, 1911; Binet & Simon, 1916). Typical items on the test required children to define common words, name objects in pictures, tell how two objects are alike, draw designs from memory, repeat back a string of spoken digits, and answer abstract questions such as "When a person has offended you and comes to offer his apologies, what should you do?"

Binet ordered his hodgepodge of tests from simple ones, which most two-year-old children could answer, to difficult ones, which children could not answer but most adults could answer. The age associated with the most difficult tasks that the child could perform was designated the child's mental age, which was then compared with the child's chronological age. In 1911 William Stern (1871–1938) proposed that mental age be divided by chronological age and then multiplied by 100 to produce the familiar IQ score. Using this formula, if a 10-year-old child is able to answer most of the items that a typical 12-year-old could answer, then the 10-year-old child's IQ score would be $(12/10) \times 100 = 120$. More recently, IQ has been measured by looking at the average for the age group and determining how far above or below the average the test taker's score lies. Average is set as equal to 100; standard deviation (a measure of dispersion) is usually set as equal to 15. Using this formula, a person who scores two standard deviations above the average would be assigned an IQ score of 130.

Binet did not believe that an IQ score was a measure of intelligence, which he regarded as too complex to capture with a single number. He made it clear that IQ was not like weight or height, in that IQ does not represent a quality possessed by a person. Again, Binet believed that his test was good only as a guide to help identify children who needed special help. Furthermore, Binet did not believe that scores on IQ tests necessarily represented a genetically based intellectual potential. Rather, he was optimistic that, with special education, many children who scored low on the IQ test could greatly improve their reasoning, memory, and verbal skills. Binet recommended that special education be tailored to the individual's needs and aptitudes, that classrooms for special education be kept small, and that the initial focus be kept on motivation and work discipline.

Correlates of IQ

Since the early 1900s, a large number of intelligence tests have been developed. These include the Stanford-Binet (a modification of Binet's original test), the Wechsler scales for children (WISC) and adults (WAIS), each of which com-

putes a verbal IQ score and a performance IQ score; the Raven's Matrices, a nonverbal test of intelligence; and college entrance tests like the SAT.

Research on IQ tests demonstrates that various IQ test scores are positively correlated with one another; for example, the Wechsler IQ score correlates about .8 with the Stanford-Binet. IQ tests are also moderately correlated with grades in school (the correlation is usually about .5), number of years of formal education, occupational status, and, to a lesser extent, with success in an occupation (Kline, 1991; Neisser, Boodoo, et al., 1996). The correlation between success in an occupation (measured, for example, by supervisor ratings) and IQ scores is typically about .3. So people who get good grades, go to school for a long time, have professional jobs such as doctors and lawyers, and get higher ratings from supervisors evaluating their work tend to score higher on IQ tests than do people who get poor grades, drop out early, have jobs such as factory workers, and get lower evaluations from their supervisors.

Keep in mind that these correlations do not tell us much about the causes of the relationship between IQ scores and other measures, such as grades in school. It could be, for example, that the superior intellect some people possess causes them to score higher on IQ tests, do better in school, and get better jobs. But there are other possibilities. Perhaps motivation to succeed is the cause (or at least one of the causes) of the correlations—a generally motivated person will try harder to do well on IQ tests, stay in school longer, and work harder on the job. Or maybe health is a cause of the correlations—a generally healthy person is more likely than an unhealthy person to be alert in school, acquire the knowledge necessary to do well on IQ tests, and perform well on the job. It could also be that the economic advantage some people enjoy is what enables them to do better on IQ tests, do better in school, and get better jobs (McClelland, 1973).

It should be pointed out, however, that the relationship between IQ performance and educational and occupational success cannot be attributed entirely to socioeconomic factors (Barrett & Depinet, 1991). Parental background variables like parental income and education do not predict occupational achievement as well as do IQ test scores (Gottfredson & Brown, 1981). Grades in school are more strongly correlated with SAT scores than with parental income (Baird, 1984).

36.2 Is Intelligence Unitary?

As I suggested at the beginning of this chapter, much of the recent work on the nature of intellectual differences has taken the form of a reaction to the historically entrenched hereditarian theory of intelligence and the IQ enterprise it established. In this section I will discuss the evidence that intelligence is unitary, that it reflects a generic intellectual system. In the next major section I will develop the argument for a multifaceted model of intellectual differences.

Evidence for the Unitary View

Charles E. Spearman (1863–1945) was one of the first psychologists to demonstrate that people who do well on any one subtest of the IQ inventory tend to do well on any other subtest. That is, the various subtests that make up the IQ

inventory are positively correlated (Kline, 1991). Spearman thought that the prevalence of positive correlations reflected a physical property of the brain, namely, a kind of mental energy that some brains happened to possess more of than other brains (Spearman, 1927). He labeled this idea "g," to stand for the *general factor* that underlies all intellectual activity. More recent but similar interpretations of g are that g reflects the capacity to pay attention to information (Hunt, 1980; Jensen, 1979), reflects nerve conduction velocity and rate of neural decay (Jensen, 1993), or reflects the ability of neurons to change connections (Larson & Saccuzzo, 1989).

Spearman, and many others since, have noted that subtests of the IQ inventory that are similar to one another are even more positively correlated than are dissimilar subtests. For example, two different subtests that measure spacial reasoning will be more highly correlated than a subtest that measures spacial reasoning and a subtest that measures vocabulary. This pattern of correlations, analyzed by a statistical technique called factor analysis, is sometimes interpreted as indicating that intelligence has a general (also known as *fluid*) component that reflects some genetically determined biological aspect of the cognitive system, and a series of specialized (also called *crystallized*) components that reflect various learned skills (Kline, 1991).

There is other evidence for the unitary nature of intelligence. Correlations among IQ tests are significant even when one IQ test is verbal and the other IQ test is nonverbal. For example, the correlations between the Raven's Matrices (a nonverbal IQ test) and conventional IQ tests range from about +.40 to +.75 (Anastasi, 1988). That IQ scores predict performance in very different situations, such as school settings and job settings, also suggests that there is a unitary aspect to intelligence.

What Underlies Unitary Intelligence? Contributions of Information Processing

Elsewhere I have criticized information processing models that postulate that all problems are solved by the same, generic information processing system. A similar sort of information processing perspective has been used as an account for why intelligence seemingly has a unitary character.

A generic information processing approach to intellectual differences has all intellectual tasks performed by a single information processing system. Individual differences in intellectual performance reflect differences in the speed and efficiency with which the various components of the system are executed. I do wish to note that information processing cognitive psychologists need not postulate a generic information processing model of individual differences. Perhaps human cognition is composed of many different, relatively autonomous information processing systems. However, the idea of a generic information processing system is implicit in most information processing approaches to cognition (Lachman, Lachman, & Butterfield, 1979), and so it is the generic form of it that I will critique here.

The information processing approach rose to prominence in the 1950s, 1960s, and 1970s. An important claim of information processing is that any given cognitive process can be broken down into a set of fundamental components, such as perceiving information, transforming information, storing information in memory, and retrieving symbols from memory. Most information processing

accounts claim that there is a limited-capacity working memory—a place that holds the currently activated information and the program for manipulating it. More discussion of information processing can be found in the introductory chapter.

The contribution of information processing to the study of intelligence is its claim that any or all of these components of cognition could be the basic and essential source of individual differences in intellectual activity (e.g., Carroll, 1983; Jensen, 1982; Vernon, 1983; Sternberg, 1985; Hunt, 1983; Pellegrino & Glaser, 1979). Some people might be more intelligent than others because they can more quickly and efficiently process stimulus input, retrieve information from memory, or transform information from one form into another.

An Example of Research Based on Information Processing: Inspection Time The information processing perspective has produced a variety of experimental paradigms for measuring the speed and efficiency with which people can carry out any component of cognitive processing. In the typical information processing experiment, researchers use established experimental paradigms to obtain from each subject an estimate of how quickly or efficiently the subject can execute one of the components, and then measure the correlation between that estimate and the subject's score on an IQ test.

One task that has been studied extensively is called the *inspection time task* (Deary & Stough, 1996). In a typical version of this task, subjects are given two parallel vertical lines joined at the top by a horizontal line. One of the vertical lines is longer than the other. An example of a stimulus used in the inspection time task is provided in figure 36.1. Over a series of trials the longer line is presented on the left side about as often as it is presented on the right. Subjects must identify which is the longer line and can take as long as they want to make the decision. The task is made difficult by limiting the amount of time the stimulus is exposed to the subjects; that is, the inspection time is kept brief. The range of exposure durations is usually between 100 milliseconds to less than around 10 milliseconds. Any given subject's inspection time is usually expressed as the stimulus duration necessary for the subject to reach a given ac-

Figure 36.1
A typical stimulus used in the inspection time task. From a very brief exposure to such a stimulus, subjects must decide whether the left or the right vertical line is longer. (See Deary and Stough, 1996.)

curacy level, such as 75%. Be clear that inspection time does not refer to how long it takes a subject to make this simple discrimination; rather, it refers to how long the stimulus was exposed in order that the subject might reach an acceptable level of performance.

The main finding of interest is that inspection times correlate with performance on standard tests of intelligence (e.g., Nettelbeck & 1976; Deary, 1993; see Deary & Stough, 1996). People whose inspection times are short tend to score higher on the intelligence tests. Across a variety of studies the correlation is usually around .5, a moderately strong correlation (Deary & Stough, 1996). One interpretation of the correlation is that inspection time measures a basic information processing component—namely, the speed with which information is taken in or initially perceived.

Other information processing measures have also been correlated to IQ. These include estimates of the span of working or short-term memory (Hunt, 1978; Schofield & Ashman, 1986; Daneman & Carpenter, 1980; Dark & Benbow, 1991; see Dempster, 1981), the speed with which subjects supposedly scan short term memory (Keating & Bobbit, 1978; Vernon, 1983), the speed with which people mentally rotate a visual stimulus (Mumaw Pellegrino, Kail, & Carter, 1984), the speed with which people access the name of a letter (Hunt, 1978, 1983; Hunt, Lunneborg & Lewis, 1975), and the speed with which subjects access the meaning of a word in memory (Goldberg, Schwartz, & Stewart, 1977; Vernon, 1983). Measures of the speed of information processing tasks correlate with scores on IQ tests even when the IQ test itself is not timed (Vernon & Kantor, 1986).

Problems with the Information Processing Perspective on Intellectual Differences
There are, however, problems with the information processing account of individual differences in cognition. One problem is that not every researcher finds a correlation between measures of the speed or efficiency of a component and IQ performance (e.g., Keating, 1982; Ruchalla, Scholt & Vogel, 1985; see Longstreth, 1984; Barrett, Eysenck, & Luching, 1989). Further, when a correlation is found, that correlation is often achieved by comparing college students to mentally retarded people. When the studies are done using subjects who are not mentally retarded, the correlation between any estimate of the speed with which a cognitive component is executed and IQ scores is usually quite modest, in the .3 to .4 range (see Kline, 1991; Mackintosh, 1986). The correlation between inspection time and IQ scores seems a bit more robust, however (Deary & Sough, 1996).

A more fundamental problem is that the information processing approach relies too much on establishing correlations between measures of information processing and IQ scores. What is generally lacking from this line of inquiry are demonstrations that measures of information processing can predict performance on real life tasks better than conventional IQ tests (Richardson, 1991).

Another difficulty with the information processing approach to individual differences is that of establishing cause and effect. Is the efficiency with which information is initially processed the cause of intelligence, or is speed of processing the effect of intelligence, whose cause is undetermined? Even if it is conceded that perception speed, as measured in tasks like the inspection time

task, is a causal determinant of intelligence, what then causes there to be differences in perception speed (see Richardson, 1991)? What would be the biological basis for mental speed, or for any other component of cognition measured by information processing tasks?

Research on the Physiological Basis of Intelligence Another way to get at the underlying nature of intelligence is to examine neurophysiological correlates of individual differences in cognition. Typically, research and theory studying the neurophysiological basis for intelligence has assumed, at least implicitly, that intelligence is a unitary phenomenon. For instance, researchers have speculated that the brain of a highly intelligent person has more synapses among neurons (Birren, Woods, & Williams, 1979), more efficiently metabolizes energy (Smith, 1984), or more efficiently reconfigures connections among neurons (Larson & Saccuzzo, 1989). Unfortunately, it is difficult to obtain clear-cut evidence for or against any of these conjectures, because the research on the physiological underpinnings of individual differences in cognition is meager and inconclusive. One of the main difficulties lies in measuring the critical physiological processes, which are likely to be dynamic phenomena reflected in the way neurons communicate with one another.

Are Smart Brains Metabolically Efficient? Recently, brain imaging technology, such as positron emission tomography scanning (PET scans), has allowed researchers to study metabolic activity in various sections of the brain of an alive and awake person. Some studies suggest that people who do better on intelligence tests tend to display lower neural metabolic activity. Haier, Siegel, Nuechterlein, et al. (1988) found that performance on the Raven's Matrices was negatively correlated with overall cortical metabolic rate. Subjects who scored higher on the Raven's Matrices test (a nonverbal intelligence test) tended to have lower overall cortical metabolic rates than subjects who scored lower on the test. The authors speculated that people who are good at reasoning tasks have more efficient neural circuits which therefore use less energy than the neural circuits of people who have more trouble with the reasoning tasks.

Haier, Siegel, MacLachlan, et al. (1992) measured cortical metabolic activity during the initial stages of learning the complex computer game TETRIS, and again several weeks later after subjects practiced the game. They found that subjects who improved the most on the computer game displayed the largest drop in cortical metabolic activity while playing the game. Similar results have been found by Parks, et al. (1988).

In apparent contradiction to these studies, though, is research that has uncovered a positive correlation between metabolic rate and performance on IQ tests. This research, however, has usually used elderly subjects, some of whom have Alzheimer's disease and other forms of dementia (e.g., Butler, Dickinson, Katholi, & Halsey, 1983; Chase et al., 1984). Aging and disease may alter the normal functioning of the brain.

Even if the negative correlation between performance on intelligence tests and cortical metabolic activity proves reliable, interpretation problems remain. It is not clear what makes neural circuits more efficient. Is it the density of the neurons, the ease with which neurons affect the activity of other neurons, the number of glial cells that support the neurons, or any of a number of other

possibilities? Furthermore, there may be other reasons for the slower cortical metabolic rate in people who score higher on the intelligence tests. Perhaps people who are able to remain calm while taking intelligence tests have lower cortical metabolic rates as a result, and thus do better on the tests. Both intelligence test performance and metabolic rate may be affected by control over anxiety. Haier et al. (1988) dismiss this possibility because their subjects did not appear to be anxious, and because other research suggests that anxiety increases metabolic rates primarily in the frontal lobes. The authors found that metabolic rate changes related to learning occurred primarily in the posterior regions. However, it is possible that anxiety responses interacting with the responses necessary to do cognitive tasks may produce a different pattern of cortical metabolic rate than observed in other situations.

Neural Conduction Rate and Smart Brains Some recent research suggests that the rate at which neurons conduct electrical activity may be faster for people who score higher on intelligence tests. Reed and Jensen (1992) presented subjects with visual stimuli and measured the latency with which an evoked potential was detected in primary visual cortex. Shorter latencies imply faster neural conduction. They found a .37 correlation between scores on the Raven's Matrices test and conduction rates. Similar findings have been reported by Vernon and Mori (1992).

Again, though, the conduction latency results are not easy to interpret. What is different about the neural structure between people whose neurons conduct impulses faster and people whose neurons conduct impulses more slowly? Does the variation in conduction latency reflect intellectual efficiency, motivation, consistency of performance, or what?

Let me make one final comment on the studies of the physiological basis of individual intellectual differences. It is possible that certain physiological features on which people differ and which determine intelligence permeate much of the brain. There may be something about the development of neurons such that virtually all of them are more efficient in some people. In such a case, intelligence would have a unitary character, as much of the research on the neurophysiology of intelligence implicitly assumes. On the other hand, it is also possible that the relative efficiency of neurons varies across neural domains within any given brain. Such variability in efficiency within a single brain could be due to environmental experiences, genetic "programming," or some interaction between the two. At any rate, neural domain variability would give rise to a multifaceted form of intelligence. And it is to a multifaceted view of intelligence that I will now direct my discussion.

36.3 Building the Case for a Multifaceted Approach to Intelligence

Interpreting the Evidence for Unitary Models of Intelligence

There have been a number of reactions to the unitary intelligence interpretation of the positive correlations observed among various measures of intelligence and information processing. One reaction is that the g factor has many possible interpretations besides the interpretation that it reflects the intrinsic efficiency of the cognitive system (Gould, 1981; Richardson, 1991). One possibility is that

g reflects the encouragement people receive as they grow up. Children who are encouraged to learn and perform well, or are made to feel secure, may try harder and/or be less anxious when taking the various subtests of the IQ test. Such children would be expected to do well on the subtests of the IQ inventory and well in academic situations. Certainly it has been established that measures of a person's attitude and motivation tend to correlate with that person's performance on IQ tests (see Anastasi, 1988). For example, people who have positive attitudes toward learning and have a desire to succeed tend to do better in school and score higher on tests of intelligence (Anastasi, 1985; Dreger, 1968).

The connection between performance in information processing paradigms and performance on IQ tests may also be interpreted as a matter of attitude and motivation, and not necessarily a matter of the intrinsic efficiency of the cognitive system. Consider that from the perspective of the subject, tasks like the inspection time task are tedious. Subjects who try hard, especially by concentrating on every trial, will tend to have short inspection times. Subjects who occasionally let their attention wander, on the other hand, will get the occasional long inspection time that will increase their overall average (Mackintosh, 1986). If the subjects who try hard on the information processing tasks are also the ones who try hard on the IQ test, then there will be correlations between measures of the information processing task and IQ performance, as is observed.

Expanding the Concept of Intelligence: Creativity, Sociability, Practicality

To some extent, the issue of whether a given test of the intellect correlates with other tests depends on what sorts of tests one wishes to consider as revealing of intelligence. When people are given tests that are dissimilar in content to those found in conventional IQ inventories, researchers often find that performance on such tests (e.g., writing plots from descriptions of short stories) does not correlate with performance on the conventional tests (Guilford, 1964, 1967; Thurstone, 1938).

Creativity One way to expand the concept of intelligence is to consider creativity as an aspect of intelligence. Recall that I first discussed creativity in chapter 8: Problem Solving. Most IQ tests have no measures of creativity, an admittedly difficult concept to define and measure objectively. Creativity usually refers to ideas or works that are novel and valuable to others. Einstein was creative when he declared that "$E = mc^2$," because the equation was novel and valuable, at least to physicists. Had he declared "$E = mc^3$" his equation would still have been novel, but not valuable.

A variety of research suggests that creativity, as measured by peer assessments, number of publications, and so on, bears little relationship to scores on IQ tests (Baird, 1982; Barron, 1969; MacKinnon, 1962; Wallach, 1976; see Perkins, 1988). For example, Yong (1994) studied Malaysian secondary students and found that a test of figural creativity was unrelated to scores on the Cattell Culture Fair test of intelligence.

Some disciplines requiring creativity tend to be populated by people who score high on IQ tests. For example, if one were to examine the general population, one would find a positive correlation between creative achievement in

architecture and IQ performance. That is because nearly all of the creative efforts are accomplished by professional architects who, as a group, do well on IQ tests. But if one examines only professional architects, one does not find a strong relationship between degree of creative achievement (measured by peer ratings of creativity) and IQ performance (MacKinnon, 1962). Similarly, among psychology graduate students, Graduate Record Exam scores did not correlate significantly with faculty advisor ratings of the students' creative abilities (Sternberg and Williams, 1997). These results suggest that people who score very low on IQ tests tend to show less evidence of creative talent than people who score higher on IQ tests. But among people whose IQ performance is in the average-to-above-average range, IQ is at best only weakly related to performance on tests of creativity. Creativity, then, is a different aspect of the intellect or involves a different kind of motivation than the skills and motivations that enable people to do well on IQ tests (McDermid, 1965; Richards, Kinney, Benet, & Merzel, 1988).

Social Skill One might also consider social skill as an aspect of intelligence, although IQ tests do not usually measure it. Again, social skill is a concept that is difficult to measure objectively. Research on social skill suggests that if social skill is measured using the same sorts of items that appear on IQ tests, then measures of social intelligence do correlate with IQ performance. An example of this is that memory for face–name associations and the tendency to correctly answer multiple choice questions about what to do in social situations are correlated with performance on IQ tests, especially IQ tests that measure verbal skills (Thorndike, 1936; Woodrow, 1939).

When social skill is assessed by directly observing people in social situations, however, there seems to be almost no relationship between it and IQ performance. Wong, Day, Maxwell, and Meara (1995) showed that people's performance on tests designed to measure cognitive aspects of social intelligence was only weakly related to their performance on tests of behavioral aspects of social intelligence. Frederiksen, Carlson, and Ward (1984) observed the interviewing skills of medical students who had to interact with "simulated" patients in several types of situations, including one in which the medical students had to inform the patient that she had breast cancer. Various aspects of the students' interviewing performance were rated by independent judges, in order to obtain a social skill score for each medical student. These scores were unrelated to the medical students' IQ scores and unrelated to their knowledge of science, as assessed by another test. Similarly, Rothstein, Paunonen, Rush, and King (1994) found that social–personality variables, especially self-confidence and a willingness to be the center of attention, predicted classroom performance (presenting convincing solutions, communicating clearly, and contributing to others' learning) in graduate school better than did standard measures of intellectual aptitude.

Practical Intelligence Most people recognize a distinction between academic intelligence (book smarts) and practical intelligence (street smarts) (Sternberg, Wagner, Williams, & Horvath, 1995). Academic intelligence as measured by standard IQ tests is disembedded from an individual's ordinary experience. Practical intelligence, however, has to do with the actual attainment of goals

that are valued. Sternberg and his colleagues (see Sternberg et al., 1995) have developed tests that supposedly measure *practical intelligence* (also known as *tacit knowledge*). Their tests typically present subjects with a set of work-related problems (e.g., how to achieve rapid promotion within a company) along with choices of strategies for solving the problem (e.g., write an article on productivity for the company newsletter, find ways to make sure that your supervisors are aware of your accomplishments). The subjects rank-order the strategies according to which is likely to achieve the goal. Their responses are then compared to those of acknowledged experts or to established rules of thumb used by experts. The greater the response overlap between subject and expert, the higher the subject's score on the test of practical intelligence.

A variety of studies suggest that scores on tests of practical intelligence correlate moderately with success on the job (see Sternberg et al., 1995). For instance, in one study, the correlation between practical intelligence test scores and performance ratings for the category "generating new business for the bank" was .56 (Wagner & Sternberg, 1985). However, scores on tests of practical intelligence are essentially unrelated to performance on standard IQ tests (Wagner & Sternberg, 1990). For instance, among Air Force recruits, the median correlation between scores on a test of practical intelligence and scores on various batteries of a standard IQ-type test was −.07 (Eddy, 1988, in Sternberg et al., 1995).

Similarly, Ceci and Liker (1986) found that, among avid racetrack patrons, the complexity of reasoning about handicapping horse races and success at predicting a horse's speed was unrelated to their IQ performance. Dorner and Kreuzig (1983) found that the sophistication of strategies used to solve city management problems was unrelated to a person's IQ. Yekovich, Walker, Ogle, and Thompson (1990) found that expertise in football, and not IQ, predicts who identifies the important facts in a passage about football, and who derives appropriate inferences about a football game. Lave (1988) showed that subjects who were easily able to perform algebraic calculations in the context of selecting which product is the best buy in a supermarket were unable to perform essentially the same calculations when the calculations were presented as math problems on a paper-and-pencil test.

The main point of the studies on creativity, social skill, and practicality is that if we expand our sense of the intellect, we find that people are not equally skilled in all areas. These observations suggest that the prevalence of g (the tendency for performance on the subtests of the IQ inventory to correlate) is largely an artifact of the restricted range of skills that the IQ inventory samples. It is probably true that the range of skills prized in academia tends to be limited to mathematical, reasoning, and verbal skills. Creativity, social skill, and practical skill, among other examples, are not usually emphasized in school.

Gardner's Frames of Mind

Howard Gardner, a cognitive scientist from Harvard University, proposed an influential theory on intelligence in a book entitled *Frames of Mind* (Gardner, 1983). In contrast to unitary theorists, Gardner postulated six distinct, relatively autonomous categories of intelligence. These categories are *verbal intelligence*, exemplified by the poet; *logical intelligence*, exemplified by the mathematician;

musical intelligence, exemplified by the composer; *spatial intelligence*, exemplified by the painter; *bodily-kinesthetic intelligence*, exemplified by the athlete or the dancer; and *social-emotional intelligence*, exemplified by the political leader or gifted parent. Gardner claimed that intellectual skill in one category is unrelated to intellectual skill in any other category. Similar claims have been made by Guilford (1964, 1967) and Thurstone (1938).

There are several remarkable features of Gardner's theory. First, he acknowledges the wide range of intellectual competencies that may be regarded as aspects of intelligence. Very few IQ tests examine the social-emotional realm, probably because, as I noted before, it is difficult to develop objective tests to see how well a person can motivate another person or understand his or her own feelings. Yet these sorts of skills are among the most prized in virtually all cultures. Very few IQ tests examine the musical or bodily-kinesthetic realm, because in our culture the intellect has historically been equated with verbal and logical intelligence. We have a hard time regarding a talented musician or dancer or athlete as unusually intelligent. Yet in many other cultures, these sorts of competencies are so regarded.

Evidence for Gardner's Frames Gardner's theory is also remarkable for the kinds of evidence used to support it. Gardner has broken with the IQ tradition of examining patterns of correlations among subtests of the IQ inventory. Instead, he uses brain damage evidence, isolated talents, anthropological evidence, and the nature of mental operations to support his theory.

The brain damage evidence suggests that damage can interfere with one intellectual competency but leave the others intact. Damage to the left frontal and temporal regions of the brain can interfere with the use of language, but leave other skills, like logical or musical skill, intact. Damage to posterior portions of the right cerebral hemisphere can produce amusia—a difficulty in expressing and appreciating music. Yet spoken language, which also uses the auditory system, is unaffected. Similarly, damage to the anterior portions of the frontal lobes can interfere with certain aspects of emotional expression, yet language and all the other intellectual skills may remain intact.

The phenomenon of isolated talents also provides evidence for Gardner's theory. There are cases of people who are unusually talented in one realm, such as music or art, but are unremarkable and sometimes even retarded in other realms, such as logical reasoning. Similar support for Gardner's theory comes from the previously discussed findings that among people who score average or above on IQ tests, musical and social skill are unrelated to IQ performance, which tends to reflect language and reasoning skills (Shuter-Dyson, 1982; Frederiksen et al., 1984). Some research also suggests that logical reasoning skills are minimally correlated with language proficiency skills, especially when the logical reasoning task uses simple vocabulary (Boyle, 1987). Research also suggests that when memory span is measured using digits, it does not correlate with language proficiency, but when memory span is measured by words in a sentence, it does correlate with language proficiency (Daneman & Carpenter, 1980; King & Just, 1991; Perfetti & Lesgold, 1977). In my own research (Guenther, 1991), I found that the rate at which people could scan their memory of sentences that varied in word length (e.g., "Lions run quickly," "Lions

jog") for a target word (e.g., "lions") was unrelated to the rate at which the same people could scan their memory of pictures of objects containing a variable number of properties (a house containing a door, window, and roof, a house containing a door and window) for some target property (e.g., a picture of a particular door).

Gardner notes that people in all cultures develop and appreciate his six proposed categories of intelligence. In all cultures virtually everyone learns something about music, movement skills such as those used in sports, spatial skills such as those used in drawing or navigating, social skills such as those used in soothing a troubled child, reasoning skills such as those underlying the exchange of goods and services, and language skills necessary to communicate. Although the IQ industry and academia implicitly claim that reasoning and language skills are of overwhelming importance, in most other cultures, including segments of our own culture outside of academia, skills such as musical and social skill are also prized.

Finally, Gardner notes that the mental operations are quite different in each category of intelligence. Language, for example, uses rules of grammar for combining symbols that bear an arbitrary relationship to ideas. Music uses rhythm and pitch to create aesthetically pleasing sounds. Logical reasoning entails comparing patterns or sequences and deriving implications, often from symbols that are quite abstract. Social intelligence involves understanding emotions and motivation. The dissimilarity among these mental operations suggests qualitative differences among categories of intellectual skill.

Criticisms of Gardner's Frames Gardner's theory is not without its critics (see Sternberg, 1990; Richardson, 1991). One complaint is that it and any multifaceted theory of individual differences fail to explain the positive correlations among subtests of IQ inventories. For example, people who do well at explaining a proverb also tend to do well on spatial, nonverbal tests. A reasonable response to this complaint is the one already discussed, namely, that conventional IQ tests sample from a limited range of possibilities. There are few, if any, objective tests of musical, social, or kinesthetic skill, few measures of creativity, few tests measuring how well people learn new information, and few tests that confront people with problems like those actually encountered in real life.

Another complaint about Gardner's theory is that it seems to divide up the human intellect in a somewhat arbitrary way. Why, for example, is there no separate category for mechanical intelligence, which Gardner subsumes under bodily-kinesthetic? Is it not possible that a person could be a skilled mechanic but not a skilled dancer or athlete? Even Gardner admits, and others have found, that within a category like spatial intelligence, people who are good at one aspect of the skill are not necessarily good at other aspects of the skill. Kosslyn, Brunn, Cave, and Wallach (1984) found that people who are good at producing accurate visual images from verbal descriptions are not necessarily the same people who are able to make rapid rotational transformations of visual images. As another example, brain damage can interfere with the grammatical aspect of language but leave the semantic aspect more or less intact.

Gardner also seems to exclude categories that might be considered types of intelligence. Why is there not a category for religious intelligence? Have not virtually all cultures developed religion? Or for culinary intelligence? Is not food preparation essential to survival and is it not related to the brain mechanisms underlying olfactory and taste perceptions? Why not a category for practical intelligence? Are not measures of practical intelligence related to performance on the job (Sternberg, Wagner, Williams, & Horvath, 1995)?

It seems, then, that there may be an inherent arbitrariness to picking categories of intelligence. The concept of intelligence seems to reflect the values and ideology of a culture, or of an institution within a culture. Different value systems imply different notions of intelligence and different ways to measure intelligence. Advocates of this intelligence-as-ideology position include Garcia (1981), Berry (1974), Heath (1983), Helms (1992), and Keating (1982). From their perspective, the notion that one possesses a single kind of intelligence may be regarded as absurd. People possess skills of varying kinds that may be measured in many ways. Actually describing a skill and inventing a way to measure it reflects the values and goals of institutions, and not some essence of intelligence residing in a person. IQ tests tend to reflect the value the academic culture places on verbal and abstract reasoning skills, and on the objective measurement of people.

I think, then, that the unitary or generic view of intelligence is misleading. Instead, intelligence is multifaceted; it reflects performance on particularized, relatively autonomous skills. As I mentioned before, the multifaceted model of intelligence is reminiscent of the domain-specific nature of problem solving (discussed in the previous chapter). Just as there is no generic problem-solving system that kicks into action whenever a problem is encountered, there is no single unitary trait that permeates all of human cognition and gives rise to individual differences in intellectual performance.

36.4 Is Intelligence Determined Primarily by Genes?

Explicit in the hereditarian theory of intelligence is the idea that intelligence is a genetically determined intellectual potential. IQ is supposed to be an approximation of the amount of this potential. In this view, then, intellectual differences among people are largely attributable to their genetic differences. Most advocates of the genetic basis for intelligence concede that the environment can either nurture or thwart the acquisition of intellectual competency. But they contend that genes are the primary determinant of one's intellectual potential, and that in most cases IQ performance provides a rough index of this potential.

The hereditarian claim is often taken to imply that: (a) environmental intervention is not likely to help people who are "intellectually at risk" and that (b) ethnic or racial differences in IQ performance are caused primarily by genetic differences, and not by social or cultural factors. It is important to see that advocates of the hereditarian theory need not draw these implications, as I will discuss later. Indeed, my main purpose in this section is to demonstrate that the evidence for a genetic component to intellectual differences does not support these two claims.

Relationship	Correlation
Identical twins reared together	.86
Identical twins reared apart	.72
Fraternal twins reared together	.60
Siblings reared together	.47
Siblings reared apart	.24
Biological parent and child, living together	.42
Biological parent and child, separated by adoption	.22
Unrelated children living together	.32
Adoptive parent and adopted child	.19

Figure 36.2
Familial correlations in IQ performance. The source of these correlations is Bouchard and McCue (1981).

Evidence for a Genetic Basis for Intelligence

Familial IQ Correlations Advocates of the hereditarian theory base the genetic hypothesis on the finding that intelligence (at least as measured by IQ tests) runs in families. For example, the correlation between parents' and children's performance on IQ tests is about .4 (see Bouchard & McCue, 1981, or Kline, 1991, for references on familial correlations in IQ scores). Especially compelling is the finding that the correlation between the IQ scores of children adopted at birth and the IQ scores of their biological parents is higher (about .32) than is the correlation between the children's IQ scores and the IQ scores of their adopted parents (about .15) (Horn, Loehlin, & Willerman, 1975). Figure 36.2 provides a table of familial correlations in IQ performance.

Evidence relevant to the genetic hypothesis comes from research on identical twins reared apart (e.g., Bouchard & McCue, 1981; Shields, 1962). In this situation, the individuals have virtually the same genes, but grow up in different environments. The usual finding is that the correlation between the IQ scores of twins reared apart is about .7, a high correlation. So despite a dissimilarity in environments, identical twins reared apart score about the same on IQ tests. This correlation is almost as high as the correlation in IQ between identical twins reared together (about .8) and much higher than the correlation between biologically unrelated siblings reared together (about .3). Biologically unrelated siblings reared together share family environments but not genes. So the inescapable conclusion seems to be that genes are a primary determinant of intelligence, at least as measured by IQ tests.

Problems with the Evidence Supporting a Genetic Basis for Intelligence The interpretation that the familial IQ correlations support an overpowering influence of genes on intelligence is problematic, however. The pattern of familial correla-

tions does not rule out a substantial influence of the environment on intellectual differences. After all, people learn child-rearing practices and other skills relevant to the cognitive development of the child from their parents. For example, children may acquire an interest in reading from their parents and pass this interest on to their own children. Children who become interested in reading are likely to read more, get better at reading, and so do well on IQ tests that are typically saturated with test items that depend on language skills.

The importance of the environment in accounting for familial IQ correlations is suggested by the fact that children and their parents are likely to grow up in similar cultural and economic circumstances. Even adopted children are likely to be placed in homes similar in educational and economic background to the homes of the biological parents. Furthermore, children adopted as infants may be more likely to have suffered from prenatal problems, which may undermine their intellectual development and reduce the correlation between their IQ scores and the IQ scores of the adopted parents.

Some of the familial correlations demonstrate an important effect of environment on intellectual differences. The IQ correlation between unrelated children living together is about .3, which is certainly much greater than zero. So there is at least some tendency for people who have dissimilar genes but similar family backgrounds to have similar IQ scores. Also, the correlation between IQ scores for ordinary biologically related siblings is about .4, which is much lower than the correlation for fraternal twins reared together (about .6), even though the genes of fraternal twins are no more similar than the genes of ordinary siblings. Presumably, though, the family environments of fraternal twins are more similar than the family environments of ordinary siblings, because twins share the same period of family history.

Turning to the twins-reared-apart paradigm, it is worth noting that the environments of twins reared apart are not necessarily all that different from those of siblings reared together. As I mentioned before, adoption agencies usually try to place adoptees in homes similar to the home of the biological parents. Furthermore, when twins are raised separately, one twin is often reared by another family member; twins are not usually separated until later childhood; and the twins often remain in contact with one another. In other words, there is a kind of environmental "contamination" that may make the environmental influences on the twins reared apart more similar than is commonly appreciated. Finally, twins are more susceptible to prenatal trauma, which can result in mental retardation, reflected in lowered IQ scores for both twins, even if reared apart. This inflates the IQ correlation between twins (see Anastasi, 1988).

Another kind of problem with the twins-reared-apart paradigm is that it does not identify which shared genes are the underlying cause of the similarity in IQ scores. One possibility is that the genes that produce the high correlations influence biological functions that are directly involved in many cognitive processes. But there are other possibilities.

Consider the following hypothetical scenario. Identical twins share facial and bodily features, the characteristics of which are established primarily by genes. How people are treated depends to some extent on their physical appearance. Consequently, people's social skills, confidence, and so on depend to some ex-

tent on their physical appearance. Social skills and confidence, in turn, may influence how one performs on IQ tests. The result would be that identical twins, even when reared apart, will tend to perform similarly on IQ tests, yet the similarity in performance has nothing to do with their intellectual potential. Instead, it has to do with their physical appearance. It could be that one twin, should she or he grow up in an environment that downplays physical appearance, might obtain a very different IQ score than the other twin.

There are other hypothetical examples I could work out. Maybe, for example, the similarity in IQ between twins reared apart is due to similarity of their metabolic rates, or to their resistance to diseases, or to any of a number of other factors that may be genetically inherited and indirectly affect performance on IQ tests. The point of these hypothetical examples is to show that establishing that twins reared apart perform similarly on IQ tests does not necessarily prove that there is a direct genetic basis for intellectual performance. Incidentally, the same argument can be made with respect to the higher IQ correlation between the biological parents and their children whom they do not raise than between the adoptive parents and those same children. Some of the genes the adopted children inherit from their biological parents influence their scores on IQ tests, but it remains unclear what aspect of biology those inherited genes control.

The Role of Environmental Factors in Intellectual Differences

One of the unfortunate implications sometimes drawn from the hereditarian theory of intelligence is that environmental factors are likely to have a rather meager effect on intelligence. Consequently, it is not worth spending money and effort trying to improve substantially the intelligence of people who might seem "intellectually at risk." Now, strictly speaking, one need not draw this implication from hereditary theory, because hereditarians concede that the environment can have an impact on intellectual development. But the problem is that an emphasis on the genetic basis of intellectual differences can blind one to the possibility that environmental factors may have a rather potent effect on intelligence. Genetically based differences lead to the idea of inevitable differences (Gould, 1981). Yet a variety of studies have demonstrated that environmental intervention can substantially improve intellectual capabilities.

Family and School Environments Some studies have looked at the behaviors of parents to see which are correlated with their children's intellectual competence. For example, the parents' use of language correlates with their children's performance on IQ tests (Hart & Risley, 1992). Child-rearing practices also correlate with the child's intellect. White (1978), for example, found that parents who reared intellectually competent children tended to do three things: first, they provided a structured, safe, and interesting physical environment for their children. Second, they spent a lot of time helping their children solve problems. Third, they established and enforced clear-cut rules, but in a loving and respectful manner.

Such studies suggest the importance of parenting styles in the acquisition of intellectual competence. The hereditarian could, however, still argue that is the favorable genes of the parents that lead them to use reasonable parenting techniques, and that the intellectual competence of their children is mainly a

consequence of inheriting these favorable genes. A better way to show that parenting styles and other environmental variables have a causal effect on the acquisition of intelligence would be to rear one group of children under one set of environmental conditions and a comparable group under a different set of conditions. Ideally, the children should be randomly assigned to the two conditions, but random assignment is obviously socially and ethically impossible.

Still, some research comes close to performing the ideal experiment. Observations of children growing up in orphanages reveals that children who receive loving affection from the caretakers will tend to average higher on IQ tests than children who do not get the affection (Skeels, 1966). Other research has provided training to a group of low-income preschool children on the intellectual skills necessary to do well in school, and has shown that such children improve their IQ performance by an average of 10 to 15 points. Unfortunately, these sorts of studies typically reveal that the gains are temporary. By the fourth grade, the average IQ performance of the group that got the training declines to the level of comparable children who did not receive the training (Bronfenbrenner, 1974; Klaus & Gray, 1968; Ramey, Campbell, & Finkelstein, 1984). However, if the intervention program is extended into the school years, evidence suggests that the intervention has a beneficial effect on IQ performance that extends beyond the first few years of school (Lazar, Darlington, Murray, Royce, & Snipper, 1982; Miller & Bizzell, 1984).

A fairly dramatic environmental effect on IQ performance was accomplished by Garber (1988), who placed a group of children who were previously labeled to be at risk for mental retardation in an extensive home enrichment program. Garber found that, by age 6, the group scored 30 points higher on an IQ test than did a control group, and even by age 14 still scored about 10 points higher than the control group. Another dramatic case is the Carolina Abecedarian Project (Campbell & Ramey, 1994). In this project, infants from low-income families were placed into intellectually enriched environments until they began school. Compared with controls, the enriched children scored higher on tests of intelligence, even 7 years after the end of the intervention.

Generational Environmental Changes: IQ Scores Are Rising One intriguing piece of evidence for an environmental influence on IQ performance is the finding that in this century there has been a steady worldwide rise in IQ scores (Flynn, 1984, 1987; see Neisser et al., 1996). The average gain has been about 3 IQ points per decade. The result is that most intelligence tests have to be periodically restandardized in order to keep the mean equal to a score of 100. So people who score 100 on an IQ test today (in 1997) would have averaged about 115 in 1947.

No one knows why IQ scores are rising. Among the proposed reasons (see Neisser et al., 1996) is the idea that the world's cultures are becoming informationally more complex, because of television, urbanization, prolonged schooling, and so on. Such complexity then produces improvements in the development of intellectual skill. Another idea is that the IQ increases are due to nutritional improvements, perhaps the same improvements that have also led to nutritionally based increases in height. Whatever the reason, it must be something in the environment that is producing the rising IQ scores. Certainly

the gene pool of the humans species cannot be changing as rapidly as IQ scores are rising. Indeed, there is no evidence that people who score higher on IQ tests are reproducing at greater rates. If anything, the evidence suggests that people who score high on intelligence tests have lower fertility rates, at least within the last century (Van Court & Bean, 1985).

In general, then, the research is consistent with the notion that environmental factors can have a large effect on the development of intellectual competency, even as measured by conventional IQ tests. The fact that people inherit genes that somehow influence intelligence, however measured, does not mean that intelligence is immutable.

Ethnic Differences in IQ Performance

Another unfortunate implication sometimes drawn from the hereditarian theory of intelligence is based on the finding that people from minority groups, such as Native Americans and African Americans, tend to score lower on IQ tests than people from majority groups such as European Americans (Herrnstein & Murray, 1994; Neisser et al., 1996). Yet IQ tests predict academic performance among minority people, suggesting that the tests are not unreasonable measures of intelligence in minority populations (Scarr-Salapatek, 1971; Oakland, 1983). The unfortunate implication sometimes drawn from these findings is that European people possess a genetically determined intellectual potential that exceeds that possessed by peoples from other parts of the world (Jensen, 1969). Some hereditarians claim that Asian people possess the most favorable genes for intellectual potential (Rushton, 1988, 1991). Such claims have historically been used to justify racial segregation and racist social and economic policies. They have also been used to discourage the spending of economic resources on the education of people from minority cultures.

Again, the implication that ethnic differences in performance on IQ tests are genetic need not be drawn from a hereditarian theory of intelligence. It is perfectly consistent with the hereditarian view that individual differences in intelligence are primarily due to genes but ethnic differences in measured intelligence are primarily due to environmental factors. I think that the consensus position is that ethnic differences in IQ performance reflect differences in cultural environments. Specifically, the cultural environment of the typical European (and in some cases, Asian) is more conducive to learning the skills that enable a person to do well on IQ tests than is the cultural environment of the typical African American or Hispanic American or Native American.

Evidence against a Genetic Basis for Ethnic IQ Differences Several lines of evidence support the claim that ethnic differences in IQ performance are a consequence of environmental and cultural factors and not a matter of genetic differences.

First of all, when children from a minority group that typically scores lower on IQ tests are raised in the same environment as children from the majority culture, the IQ scores of those minority children are similar to the IQ scores of the majority children. Scarr and Weinberg (1976, 1983) examined the IQ scores of African-American children born of mostly lower income parents but adopted by European-American families from mostly the middle and upper middle economic brackets. The IQ scores of the adopted African Americans averaged

about 20 points higher than the IQ scores of other African Americans living in lower income circumstances. Clearly, the family environment had a huge effect on the development of skills underlying the performance on IQ tests. Furthermore, these and other adoption studies indicate that when African-American children are adopted into European-American families, their average IQ performance typically comes to be nearly equal to that of the European Americans (Flynn, 1980; Eyferth, 1961; Tizard, Cooperman, Joseph, & Tizard, 1972; Scarr & Weinberg, 1976, 1983). Yet the IQ correlation between the African-American adopted children and their biological parents is greater than the correlation between the African-American children and their adopted parents. Again, that seemingly paradoxical result is because correlation reflects rank order. The IQ scores of the adopted African-American children may have been improved by their environment, but the environment did not affect their rank order on the IQ test. The rank order of the IQ scores of the adopted children continued to reflect the rank order of the IQ scores of their biological parents.

Other research that examines children in similar environments but with different racial backgrounds has also contradicted the hereditarian claim of a racial difference in intelligence. Loehlin, Lindzey, and Spuhler (1975) examined the IQ scores of children born to German mothers and American fathers stationed in Germany after World War II. One group of children was fathered by African Americans, while the other group was fathered by European Americans. Both groups were raised by German mothers in roughly similar economic circumstances. The averages of the IQ scores of the two groups of children were equal, even though one group received half of its genes from people of African descent. Furthermore, there is no correlation between degree of African ancestry of African Americans and their performance on IQ tests (Scarr, Pakstis, Katz, & Barker, 1977).

It is true that some Asian-American people, such as Japanese Americans, score higher on average on IQ tests than do European Americans. But cross-cultural studies that take into account cultural factors, such as the proportions of rural and urban dwellers, suggest no difference between Asians and Europeans in IQ test performance (Stevenson et al., 1985). Furthermore, some Asian groups that have immigrated to the West and subsequently endured poverty in the West score lower on IQ tests than do Europeans (see Mackintosh, 1986).

Sometimes people use the high correlations in IQ performance between twins reared apart as evidence that ethnic differences in IQ performance must be due to genetic differences. In fact, though, even if one overlooks the interpretation problems associated with this paradigm, the twin findings are perfectly consistent with an environmental explanation for group differences in IQ performance.

To see why, consider the following hypothetical situation. Suppose we have three sets of twins (Jerry and Gerry, Robin and Robyn, and Sara and Seri) who are reared apart. On IQ tests, Jerry and Gerry both obtain 100, Robin and Robyn both obtain 110, and Sara and Seri both obtain 120. So the correlation between the IQ scores of the twins is 1.0. Now suppose the second member of each pair (Gerry, Robyn, and Seri) is each given extensive training so that each improves his or her IQ performance by 20 points. So now the IQ scores will be

100 and 120 for Jerry and Gerry respectively, 110 and 130 for Robin and Robyn respectively, and 120 and 140 for Sara and Seri respectively. Yet the correlation between the IQ scores of the twins will still be 1.0, because correlations reflect rank order, which remains the same. This hypothetical example makes clear that even when the correlation between twins reared apart is as high as it can be (1.0), the environment can still dramatically affect group differences in IQ performance.

Why Are There Ethnic Differences in IQ Performance? If differences in IQ performance among ethnic groups are not due to genetics, what are they due to? Nobody knows for sure (Neisser et al., 1996). A clue comes from the finding that many politically and economically disadvantaged groups from all over the world tend to do less well in school and to score lower on IQ-type tests than do the more advantaged groups (Ogbu, 1978, 1994). The kinds of minority groups that score lower on IQ tests are those that became a minority group involuntarily or those that are regarded by the culture as caste-like (Ogbu, 1978, 1994). Immigrants who come to a country voluntarily may be optimistic that they can control and improve their conditions. These groups typically do well on IQ tests. Groups that are involuntarily displaced, such as Native Americans, African Americans, and the Maori in New Zealand, or are excluded, like the "untouchables" of India or non-European Jews of Israel, may lack the conviction that hard schoolwork and serious commitment to the educational enterprise will be rewarded. It is these groups that tend to do poorly on IQ tests.

Furthermore, IQ tests take place in settings in which motivation and attitudes can affect performance (see Helms, 1992; Miller-Jones, 1989). Children from a minority culture that emphasizes the interpersonal nature of learning may be more likely to regard a lack of feedback from the tester as evidence that they are doing well on the test (Miller-Jones, 1989). These children may refrain from varying their strategies in the course of taking the test and consequently obtain a lower score. In some cultures, it is unusual for an adult who already knows the answer to a question to ask that question of a child, or for children to explain what they know (Heath, 1989; Rogoff & Morelli, 1989).

It is frequently observed that people from other cultures often misunderstand the instructions and fail to take seriously the test's requirements. Sinha (1983), for example, has provided an analysis of some of the cultural reasons why Asiatic Indians who have not been enculturated by the West have trouble with IQ tests. Asiatic Indians typically do not know that responses like "I don't know" or "I can't decide" will cause one to get lower scores on IQ tests. Also, Asiatic Indians are typically inhibited in responding, especially when the task seems pointless to them. In some cultures, such as the culture in which many African Americans are raised, a premium is placed on the creativity of responses. Sometimes African-American children are surprised to learn that they are expected to provide obvious answers on IQ tests (Heath, 1989; Helms, 1992). The creative answers they often do provide get them lower scores. Boykin (1994) argues that many African Americans are alienated from education and the accompanying psychometric enterprise because these institutions implicitly conflict with a heritage that emphasizes spirituality, harmony, expressive individualism, communalism, and orality, and not talent sorting and

talent assessment. A consequence of that alienation may be a poorer average performance on IQ tests.

Cultural Differences in Prized Intellectual Competencies

In general, then, research shows that when the cultural and economic environments of ethnic groups are roughly equated, performance on IQ tests is roughly equated as well. But impoverished minority groups involuntarily displaced or shunned by the culture as a whole tend to do poorly on IQ tests. It would be a mistake, though, to conclude from the research that the poverty and cultural alienation endured by many minorities invariably suppresses intellectual development. Rather, people from different cultures place emphasis on different kinds of intellectual development (Garcia, 1981; Heath, 1983, 1989; Helms, 1992; Miller-Jones, 1989).

IQ tests were developed by middle- and upper-middle-class Europeans and people of European descent, so it is unsurprising that the intellectual skills relevant to IQ testing are emphasized more in their culture than in most other cultures. But the skills developed in other cultures in response to their environments, including impoverished environments, may be "invisible" to IQ tests. If care is taken to develop tests that reflect the intellectual competencies prized by a minority culture, but not necessarily by the majority culture, then the minority culture will do as well, and sometimes even better, on such tests.

Heath (1983) studied children from low-income African-American families, low-income European-American families, and middle-income European-American families. She noted that, on average, there were differences in the kinds of intellectual competencies with which these children began school. The African-American children from low-income families tended to be very skilled at responding to novel situations, defending themselves against a verbal insult, and telling creative stories. The European-American children from middle-income families were typically good at responding to requests, responding quickly when timed by a psychologist administering a test, and answering "why" questions. In general, then, this study makes the point that poverty or lack of formal education does not necessarily depress intellectual development; rather, it can lead to the development of intellectual skills different from those at which well-educated Europeans tend to excel and to measure with IQ tests.

Similar conclusions may be drawn from cross-cultural studies. Berry (1974) found that people from hunting cultures tend to do better on tests of perceptual discrimination and spacial processing than people from cultures in which hunting is less important. Rice farmers from Liberia are better than Americans at estimating quantities (Gay & Cole, 1967).

Children from Botswana, accustomed to storytelling, are better than American children at remembering stories (Dube, 1982). In one of my favorite examples, Cole, Gay, Glick, and Sharp (1971) asked adult Kpelle tribespeople to sort 20 familiar objects, such as knives, oranges, and so on, into groups of things that belong together. The Kpelle separated the objects into functional groups (e.g., knife with orange) and not taxonomic groups (e.g., knife with fork). Western adults, on the other hand, sort on the basis of taxonomy, as do children who receive higher IQ scores. But when the Kpelle adults were asked to sort the objects the way a "stupid" person would do it, the Kpelle sorted like

TASKS FAVORING WOMEN

Perceptual Speed
Find the house that exactly matches the one on the left.

Verbal Fluency
Indicate another word that begins with the same letter, not included in the list.

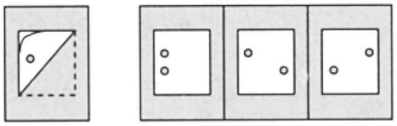
L – – –

Limp, Livery, Love, Laser, Liquid, Low, Like, Lag, Live, Lug, Light, Lift, Liver, Lime, Leg, Load, Lap, Lucid

Answers:
The house at the far right; Life or any other word beginning with L.

TASKS FAVORING MEN

Spatial Relations
A hole has been punched in the folded sheet. How will the sheet appear when unfolded?



Mathematical Reasoning
In the space at the left, write the answer to the following problem.

If only 60% of seedlings will survive, how many must be planted to obtain 660 trees?

Answers:
The middle sheet; 1,100 seedlings.

Figure 36.3
Problem-solving tasks favoring women and problem-solving tasks favoring men. From *Psychology* by Fernald, Dodge, © 1994. Reprinted with permission of Prentice-Hall, Inc., Upper Saddle River, NJ.

the Western adults and high IQ children—that is, on the basis of taxonomy. At least with respect to those objects, the typical Kpelle adult regarded the functional grouping as more useful than the taxonomic grouping.

Sex Differences in Intellectual Competencies
Perhaps because people are fascinated by male–female differences, there have been many studies of sex differences in cognition. Many of these studies report that males tend to do better in tests of mathematical and spatial ability, and females tend to do better in tests of verbal ability (reviewed in Maccoby and Jacklin, 1974; Bjorklund, 1995; Kimura, 1992; Halpern, 1992). Examples of tasks that favor males and tasks that favor females are provided in figure 36.3. Men and women do not differ in IQ scores, vocabulary tests, or reasoning tasks.

The nature of the sex differences depends on how cognitive skills are measured. To illustrate, males do slightly better than females on spatial tests that measure the ability to orient oneself in relationship to objects or to mentally transform spatial information. But females do slightly better than males on spatial tests measuring ability to learn and remember spatial relationships (Silverman & Eals, 1992). Although males do better on most objective tests of mathematical ability, females get better grades in math courses than do males (Kimball, 1989).

It should be noted that there is considerable controversy surrounding sex differences in cognition. Some researchers claim that the average differences between males and females are usually small and often statistically insignificant (Hyde, 1981) whereas others claim that the differences are substantial (Eagly, 1995). Some researchers claim that the differences may have been declining in recent years (Feingold, 1988; Voyer, Voyer, & Bryden, 1995) but others claim that the differences have remained stable (Halpern, 1992). And, of course, the biggest controversy has to do with whether cognitive differences between the sexes are due to the different genes that the sexes inherit or to the different environments and cultures in which they grow up.

Genetic Basis of Sex Differences in Cognition What is the cause of the sex differences in cognition? Obviously, boys and girls are treated differently and encouraged in different ways (Halpern, 1992). Boys are more likely to be encouraged to pursue careers in science, engineering, and mechanics, where mathematical and spatial skills are important. Girls are more likely to be encouraged to pursue careers in teaching and in child rearing, where communication skills are important.

Still, many researchers have proposed genetically based biological explanations for male–female differences in cognition (e.g., Kimura, 1992). Usually the ultimate cause of sex differences is attributed to the supposedly different selective pressures on males and females as humans evolved. Supposedly, males did the hunting, and so evolved better spatial skills for orienting to and transforming spatial information; and females did the gathering and child rearing, and so evolved better spatial memory and verbal skills.

What biological mechanism might be controlled by the genes that underlie sex differences in cognition? One example of a biological mechanism that may plausibly be coded for in the genes and that may give rise to sex differences in cognition is the production of sex hormones. Sex hormones, such as testosterone, are known to influence the organization of the mammalian brain during critical periods in prenatal development (Geschwind & Galabura, 1987; Halpern & Cass, 1994). A variety of research supports a correlation between sex hormones and performance on sex-differentiating cognitive tasks.

Women who were exposed to abnormally high levels of the male hormone androgen *in utero* score higher than do controls on tests of spatial ability (Resnick, Berenbaum, Gottesman, & Bouchard, 1986). Older males given testosterone improve on visual-spatial tasks (Janowsky, Oviatt, & Orwoll, 1994). Women do better on cognitive tasks that favor women over men, like verbal skills, and worse on cognitive tasks that favor men, like spatial rotation, when they are in the midluteal phase of the menstrual cycle than when they are in the late menstrual phase. Levels of estrogen and progesterone are higher during the midluteal phase (Hampson & Kimura, 1988; Hampson, 1990a, 1990b; see Kimura & Hampson, 1994). Men do better on tasks that favor men over women during the spring, when their testosterone levels are relatively low, than in the autumn, when their testosterone levels are relatively high (Kimura & Toussaint, 1991; see Kimura & Hampson, 1994). And it isn't just that men do better in the spring, when a young man's fancy supposedly turns to love–men's performance on tasks that do not favor men over women, such as reasoning, is the

same in spring as in autumn. Apparently, average to below-average levels of testosterone are associated with optimal performance on visual-spatial tasks in men (Gouchie & Kimura, 1991).

What is it that sex hormones do to the brain that gives rise to differences in cognition? One possibility is that hormones affect how the cerebral hemispheres distribute their function. Recall that in most people the left hemisphere is more involved than the right in the control of language whereas the right hemisphere is more involved than the left in the control of spatial processing. Perhaps the female advantage for some verbal skills reflects the involvement of more right-hemisphere neural tissue in language–neural tissue that at the same time encroaches on the neural tissue that would have been used for spatial processing. At least some evidence suggests that there is less tendency among women for their left hemisphere to control language more than their right hemisphere (e.g., Shaywitz et al., 1995), although not all studies find a sex difference in hemispheric specialization (Newcombe & Bandura, 1983; Waber, Mann, Merola, & Moylan, 1985).

Another possible neurological model of sex hormone differences in brain organization has been developed by Kimura (1992). Kimura suggests that the organization of functions within the left hemisphere differs between the sexes. For language functions, women may make more use of the anterior portions of the left hemisphere whereas men make more use of the posterior left hemisphere. Such a difference may give rise to the tendency for women to do better on tests of verbal fluency, because the grammatical aspect of language may be more anatomically connected to the planning and strategic components of information processing. The more intimate connection in males between language centers and the centers involved in visual perception may give rise to the male advantages on spatial reasoning tasks. One line of evidence consistent with this view is that aphasia (language disturbance) occurs more often in women when the damage is near the front of the left hemisphere, but more often in men when the damage is in the posterior area of the left hemisphere (Kimura, 1992).

It is important to point out that the supposed differences in the brains of men and women may not necessarily reflect the effects of sex hormones; those differences may be mediated by some other biological mechanism. Furthermore, the sex differences in relevant biological mechanisms need not be entirely or even at all due to genes. It may be that experiences, like playing with toys or studying mathematics, affect the production of hormones (and any other relevant biological mechanism) and thereby produce sex differences in certain cognitive skills.

Environmental Explanations of Sex Differences in Cognition My own belief is that it remains a viable possibility that sex differences in cognition are due mostly to environmental factors (how is that for a hedge!). One line of evidence for an environmental explanation of sex differences is that parental attitudes and expectations are correlated with performance on math (Raymond & Benbow, 1986) and verbal tests (Roe, Drivas, Karagellis, & Roe, 1985). An especially compelling line of evidence is research that shows that, with practice and feedback, women improve as much as men do on spatial tasks (e.g., Law, Pellegrino, & Hunt, 1993; see Halpern, 1992 for a review). Some cross-cultural

work shows that among Canadian Eskimos, a culture in which both males and females travel far from home and hunt, there are no differences in spatial abilities between males and females (Berry, 1966).

Indeed, at present in our culture, it is at least debatable whether there are any reliable male–female differences in verbal and math skills. Hyde and Linn (1986, 1988) reviewed 165 studies of verbal ability representing over 1.4 million people and found no average difference between males and females. Moreover, Hyde, Fennema, and Lamon (1990) reviewed 100 studies of mathematical performance and found that sex differences were quite small, but tended to favor females in large samples that are taken from the general population. It is only in the population of mathematically gifted individuals that males outperform females, on the average.

Carol Tavris, in her splendid book *The Mismeasure of Woman* (Tavris, 1992), reviews evidence that suggests that male and female brains learn, reason, and process information in similar ways. Tavris also discusses the bias against publishing research that finds no sex differences in cognition, and the unfortunate consequences this bias has for women. For example, a belief that males have superior mathematical skills, sustained by a bias against publishing studies that show no sex differences in mathematical skill, provides a rationale for excluding women from the sciences and for denigrating the few women who do manage to become scientists.

Conclusions about the Genetic Basis of Intelligence

There seems to be no easy way to summarize the evidence relevant to the genetic basis for intelligence. Because we are unable to conduct controlled experiments that vary genes and environments, we remain ignorant of how to interpret correlations in the IQ scores of individuals who share genes. Individuals who share genes almost always share environments. With regard to sex differences in cognition, it is difficult to disentangle the influence of sex-linked genes and sex-linked environments. It is true that the twins-reared-apart studies, as well as other research on adoption, suggest that something that is genetically inherited causes differences in scores on IQ tests. However, it is not clear what genetically controlled biological mechanism is responsible for the similarity in IQ scores. Indeed, at this point we do not really know what biological mechanisms are the underlying basis for individual differences in any of the potentially limitless kinds of skills a person can acquire. All we can say with certainty is that the biological mechanisms underlying intellectual development are, especially in our species, designed to enable us to learn from the environment. Consequently, any act of the intellect will invariably reflect both biological and environmental factors. Genetic models of intellectual differences to date lack any clear explanation of what biological mechanisms underlie individual differences. Sex hormones may be a basis for male–female differences in cognition; however, it is possible that sex hormone production may be the effect of different environments and not necessarily the direct cause of cognitive differences.

Any useful model needs to explain how a genetically determined biological mechanism interacts with various aspects of the environment to produce intellectual development. It seems pointless to argue about whether intellectual

development is primarily determined by the genes or by the environment, because either can dominate depending on the circumstances. If people are given no exposure to music, for example, they will not develop musical skill. If people are born deaf as a result of a genetic defect, they will not develop any musical skill.

And, of course, the role of genes and the biological mechanisms controlling intellectual differences is invariably complicated by the difficulty in defining and measuring intelligence. As I suggested in earlier sections of this chapter, a good case can be made that there are a potentially vast number of relatively autonomous skills that a person can acquire, any one of which could be assessed in many different ways. The effects of genetically controlled biological mechanisms and environmental variables could be quite different depending on what aspect of intelligence one cares to study.

My own sense is that the influence of genes and environmental variables is so complex and intertwined, the research limitations on the effects of genes so intractable, and the notion of intelligence so potentially multifaceted, that it is not possible to know exactly how genes and environmental variables interact to produce individual differences in cognition. This need not be a distressing state of affairs, however. Our goal as psychologists and educators should be to try to create the best possible environments for fostering the acquisition of intellectual competence in our children, regardless of their genetic makeup.

Summary and Conclusions

The integrating theme for this chapter was a contrast between a hereditarian approach to individual differences in intelligence and a multi-faceted approach. The hereditarian approach make two essential claims: intelligence is unitary and is determined primarily by the genes one inherits. The multi-faceted approach claims that there are many different and relatively autonomous domains of intelligence. Intellectual skill in one domain is typically unrelated to intellectual skill in other domains.

In the first section, I discussed the rise of the hereditarian approach to intelligence and the intelligence testing movement. Probably the most historically significant event in the history of intelligence testing was the development of IQ tests. IQ tests are known to be moderately correlated with grades in school, occupational status, and success in an occupation.

In section 36.2, I discussed the main evidence for a unitary view of intelligence, which is that performance on the subtests that make up the IQ inventory and between IQ scores and academic achievement are positively correlated. A generic information processing perspective proposes that intellectual tasks are performed by a common information processing system. Differences in intellectual capability are due to the speed and efficiency with which various stages of the system are executed. One line of evidence in support of the information processing perspective comes from research that shows that the shorter the stimulus exposure time at which people can accurately discriminate between the length of two lines, the higher the person's IQ score. In a sense, the rise of the information processing analysis of individual differences represents a reemergence of the ideas of Francis Galton, who espoused them about 100 years ago.

Recent physiological research has suggested correlations between performance on intelligence tests and physiological measures such as cortical metabolic rate or neural conduction speed. Usually, these neurophysiologically based models implicitly suppose that intelligence is unitary—that some aspect of neurophysiology that permeates all intellectual tasks is the factor that gives rise to individual differences in cognition. While intriguing, such research has not yet elucidated the underlying biological mechanisms or the causes of such correlations.

At any rate, correlations among IQ subtests or between IQ tests and academic success can be explained without supposing that all intellectual differences represent differences in a single underlying substrate of the various cognitive systems. In section 36.3 I discuss how the correlations could reflect motivation or the limited range of skills measured by IQ tests and taught in schools. Indeed, if one examines creativity, social skills, or practical skills used in everyday life, the correlations between such skills and IQ tests are essentially nonexistent.

One alternative to the unitary model is the claim that there are several distinct, relatively autonomous categories of intelligence. Howard Gardner (1983), for example, claims that there are six different categories of intelligence, and cites physiological and anthropological evidence to bolster his claim. Another alternative claims that there are potentially an unlimited number of categories of intelligence, any one of which may be measured in a potentially unlimited number of ways. The ways a culture defines and measures intelligence reflect the values and goals of the culture, and not something intrinsic to the biology of people.

In section 36.4 I discuss the hereditarian claim that intelligence is largely genetically determined. The claim is supported by familial correlations in IQ performance, and by the high correlation between the IQ scores of identical twins reared apart. However, the familial pattern of correlations is also consistent with a substantial impact of environmental factors on intelligence. The twins-reared-apart findings only show that some genetically determined biological mechanism underlies IQ performance. That mechanism might control intellectual processes, but it might also control physical appearance, metabolic rate, resistance to disease and/or any of a number of other traits.

One unfortunate implication sometimes drawn from a theory that emphasizes the genetic basis of intelligence is that environmental factors are likely to have minimal influence on intellectual development. In fact, though, a variety of studies demonstrate that appropriate environmental intervention can improve the intellectual performance of individuals who might otherwise be at "intellectual risk." Furthermore, performance on IQ tests is rising about 3 IQ points a decade all around the world.

Another unfortunate implication historically drawn by hereditarians is that ethnic differences in IQ performance reflect genetic differences among racial and other ethnic groups. However, adoption studies and other research convincingly makes the case that differences among the average IQ scores of ethnic groups reflect environmental and cultural differences among groups. If members from two different ethnic groups are raised in similar circumstances, their

average IQ performances will be similar as well. Furthermore, some research suggests that different ethnic groups, in response to their respective environments, are likely to develop different skills, not all of which are measured by IQ tests.

Sex differences in cognition have been explored as well. Some hereditarian approaches have claimed that the superior performance of the average male on spacial and mathematical tests, and the superior performance of the average female on verbal tests, reflect sex-linked genetic differences between the sexes. Fluctuations in sex hormones are correlated with performance on just those tasks on which the sexes are different. However, once again, the sex differences may be largely attributable to environmental factors. I am personally impressed with the research that shows that, with practice and feedback, women improve as much as men do on spatial tasks. There is some admittedly controversial evidence that sex differences in cognition are shrinking over time, possibly because of cultural changes made in recent years whereby more women are encouraged to attend college and pursue careers in which mathematical and spatial skills are important.

Certainly both the biological mechanisms put into place by the genes and the environment invariably contribute to intellectual growth and individual differences. How could it be otherwise? A useful model of biology's role in intelligence must specify precisely how any given biological mechanism responds to the various aspects of the environment in the course of intellectual development. Given that controlled experiments are ethically and biologically impossible, we may never completely understand the precise contributions that genes and environmental factors make to individual intellectual differences.

Recommended Readings

Gould's (1981) *The Mismeasure of Man* is a masterful and highly critical history of the rise of the intelligence testing movement. An equally masterful companion piece is Tavris's (1992) *The Mismeasure of Woman*, in which Tavris discusses her thesis that male–female differences in human emotions and cognition are greatly exaggerated. The case for important sex differences in human cognition is provided in an interesting *Scientific American* article by Kimura (1992). Gardner published his six categories of intelligence theory in his (1983) *Frames of Mind*, an exciting and wide-ranging book that has become highly influential in educational circles. Certainly people interested in intelligence testing should read Hernstein and Murray's (1994) best-seller *The Bell Curve*, but please read along with it reviews of *The Bell Curve* written by experts in the field; a collection of such reviews can be found in *The Bell Curve Wars*, edited by Fraser (1995). A summary of what psychologists know and don't know about intelligence and intelligence testing can be found in a recent *American Psychologist* review paper by Neisser et al. (1996).

References

Anastasi, A. (1985). Reciprocal relations between cognitive and affective development: With implications for sex differences. In T. B. Sonderegger (Ed.), *Psychology and gender* (Nebraska Symposium on Motivation, Vol. 32, pp. 1–35). Lincoln: University of Nebraska Press.

Anastasi, A. (1988). *Psychological testing*. New York: Macmillan.

Baird, L. L. (1982). *The role of academic ability in high-level accomplishment and general success* (College Board Repl No. 82–6). New York: College Entrance Examination Board.

Baird, L. L. (1984). Relationships between ability, college attendance, and family income. *Research in Higher Education, 21*, 373–395.

Barrett, G. V., & Depinet, R. L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist, 46*, 1012–1024.

Barrett, P., Eysenck, H. J., & Luching, S. (1989). Reaction time and intelligence: A replicated study. *Intelligence, 10*, 9–40.

Berry, J. W. (1996). Temne and Eskimo perceptual skill. *International Journal of Psychology, 1*, 207–229.

Berry, J. W. (1974). Radical cultural relativism and the concept of intelligence. In J. W. Berry & P. R. Dasen (Eds.), *Culture and cognition: Readings in cross-cultural psychology* (pp. 225–229). London: Methuen.

Binet, A. (1911). *Les indees modernes sur les enfants*. Paris: Flamarion.

Binet, A., & Simon, T. (1916). *The intelligence of the feeble-minded*. Baltimore: Williams and Wilkins.

Birren, J. E., Woods, A. M., & Williams, M. V. (1979). Speed of behavior as an indicator of age changes and the integrity of the nervous system. In F. Hoffmeister & C. Muller (Eds.), *Brain function in old age*. New York: Springer-Verlag.

Bjorklund, D. F. (1995). *Children's thinking: Developmental function and individual differences*. Pacific Grove, CA: Brooks/Cole.

Boring, E. G. (1950). *A history of experimental psychology*. New York: Appleton-Century-Crofts.

Bouchard, T. J., & McCue, M. (1981). Familial studies of intelligence: A review. *Science, 212*, 1055–1059.

Boykin, A. W. (1994). Harvesting talent and culture: African-American children and educational reform. In R. Rossi (Ed.), *Schools and students at risk*, (pp. 116–138). New York: Teachers College Press.

Boyle, J. P. (1987). Intelligence, reasoning, and language proficiency. *The Modern Language Journal, 71*, 277–288.

Bronfenbrenner, U. (1974). *Is early intervention effective? A report on longitudinal evaluations of preschool programs* (Vol. 2). Washington, DC: Department of Health, Education, and Welfare, Office of Child Development.

Butler, M. S., Dickinson, W. A., Katholi, C., & Halsey, J. H. (1983). The comparative effects of organic brain disease on cerebral blood flow and measured intelligence. *Annals of Neurology, 13*, 155–159.

Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study on children from low-income families. *Child Development, 65*, 684–698.

Carroll, J. B. (1983). Individual differences in cognitive abilities. In S. H. Irvine & J. W. Berry (Eds.), *Human assessment and cultural factors*. New York: Plenum.

Ceci, S. J., & Liker, J. K. (1986). A day at the races: A study of IQ, expertise, and cognitive complexity. *Journal of Experimental Psychology: General, 115*, 255–266.

Chase, T. N., Fedio, P., Foster, N. L., Brooks, R., DiChiro, G., & Mansi, L. (1984). Wechsler Adult Intelligence Scale Performance. Cortical location by flurodeoxyglucose F 18-positron emission tomography. *Archives of Neurology, 41*, 1244–1247.

Cole, M., Gay, J., Glick, J., & Sharp, D. W. (1971). *The cultural context of learning and thinking*. New York: Basic Books.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 9*, 450–466.

Dark, V. J., & Benbow, C. P. (1991). Differential enhancement of working memory with mathematical versus versus verbal precocity. *Journal of Educational Psychology, 83*, 48–60.

Deary, I. J. (1993). Inspection time and WAIS-R IQ subtypes: A confirmatory factor analysis study. *Intelligence, 17*, 223–236.

Deary, I. J., & Stough, C. (1996). Intelligence and inspection time: Achievements, prospects, problems. *American Psychologist, 51*, 599–608.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89*, 63–100.

Dorner, D., & Kreuzig, H. (1983). Problelosefahigkeit and intelligenz. *Psychologische Rundhaus, 34*, 185–192.

Dreger, R. M. (1968). General temperament and personality factors related to intellectual performances. *Journal of Genetic Psychology, 113*, 275–293.

Dube, E. F. (1982). Literacy, cultural familiarity, and "intelligence" as determinants of story recall. In U. Neisser (Ed.), *Memory observed: Remembering in natural contexts* (pp. 274–292). New York: Freeman.

Eagly, A. H. (1995). The science and politics of comparing women and men. *American Psychologist, 50*, 145–158.

Eddy, D. M. (1988). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge, England: Cambridge University Press.

Eyferth, K. (1961). Leistungen verschiedener Gruppen con Besatzungskindern in Hamburg-Weschler Intelligenztest fur Kinder (HAWIK). *Archiv fur die gesamte Pscyhologie, 113*, 223–241.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 42*, 95–103.

Flynn, J. R. (1980). *Race, IQ and Jensen*. London: Routledge & Kegan Paul.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932–1978. *Psychological Bulletin, 95*, 29–51.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 95*, 29–51.

Frederiksen, N., Carlson, S., & Ward, W. C. (1984). The place of social intelligence in a taxonomy of cognitive abilities. *Intelligence, 8*, 315–337.

Galton, F. (1883). *Inquiries into human faculty and its development*. London: Macmillan.

Galton, F. (1884). *Hereditary genius*. New York: D. Appleton.

Garber, H. L. (1988). *The Milwaukee Project: Preventing mental retardation in children at risk*. Washington, DC: American Association of Mental Retardation.

Garcia, J. (1981). The logic and limits of mental aptitude testing. *American Psychologist, 36*, 1172–1180.

Gardner, H. (1983). *Frames of mind*. New York: Basic Books.

Gay, J., & Cole, M. (1967). *The new mathematics and old culture: A study of learning among the Kpelle of Liberia*. New York: Holt, Rinehart & Winston.

Geschwind, N., & Galabura, A. M. (1987). *Cerebral lateralization: Biological mechanisms, associations, and pathology*. Cambridge, MA: MIT Press.

Gilbert, J. A. (1894). Researches on the mental and physical development of school children. *Studies from the Yale Psychological Laboratory, 2*, 40–100.

Goldberg, R. A., Schwartz, S., & Stewart, M. (1977). Individual differences in cognitive processes. *Journal of Educational Psychology, 66*, 325–332.

Gottfredson, L. S., & Brown, V. C. (1981). Occupational differences among white men in the first decade after high school. *Journal of Vocational Behavior, 19*, 251–289.

Gouchie, C., & Kimura, D. (1991). The relationship between testosterone levels and cognitive ability patterns. *Psychoneuroendocrinology, 16*, 323–344.

Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.

Guenther, R. K. (1991). Generic versus specialized information processing. *American Journal of Psychology, 104*, 193–209.

Guilford, J. P. (1964). Zero correlations among tests of intellectual abilities. *Psychological Bulletin, 61*, 401–404.

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Haier, R. J., Siegel, B. V., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., Browning, H. L., & Buchsbaum, M. S. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studies with positron emission tomography. *Intelligence, 12*, 199–217.

Haier, R. J., Siegel, B. V., MacLachlan, A., Soderling, E., Lottenberg, S., & Buchsbaum, M. S. (1992). Regional glucose metabolic changes after learning a complex, visio-spatial-motor task: A positron emission tomography study. *Brain Research, 570*, 134–143.

Halpern, D. F. (1992). *Sex differences in cognitive abilities* (2d ed.). Hillsdale, NJ: Erlbaum.

Halpern, D. F. & Cass, M. (1994). Laterality, sexual orientation, and immune system functioning: Is there a relationship? *International Journal of Neuroscience, 77*, 167–180.

Hampson, E. (1990). Variations in sex-related cognitive abilities across the menstrual cycle. *Brain and Cognition, 14*, 26–43.

Hampson, E. (1990). Estrogen-related variations in human spatial and articulatory motor skills. *Psychoneuoendocrinology, 15*, 97–111.

Hampson, E., & Kimura, D. (1988). Reciprocal effects of hormonal fluctuations on human motor and perceptual-spatial skills. *Behavioral Neuroscience, 102*, 456–459.

Hart, B., & Risley, T. R. (1992). American parenting of language-learning children: Persisting differences in family-child interaction observed in natural home environments. *Developmental Psychology, 28*, 1096–1105.

Heath, S. B. (1983). *Ways with words*. Cambridge, England: Cambridge University Press.

Heath, S. B. (1989). Oral and literate traditions among Black Americans living in poverty. *American Psychologist, 44*, 367–373.

Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist, 47*, 1083–1101.

Hergenhahn, B. R. (1986). *An introduction to the history of psychology*. Belmont, CA: Wadsworth.

Hernstein, R., & Murray, C. (1994). *The bell curve*. New York: Free Press.

Horn, J. L., Loehlin, J., & Willerman, L. (1975). Preliminary report of Texas adoption project. In Munsinger, H., The adopted child's IQ: A critical review. *Psychological Bulletin, 82*, 623–659.

Hunt, E. (1978). Mechanisms of verbal ability. *Psychological Review, 85*, 199–230.

Hunt, E. (1980). Intelligence as an information-processing concept. *British Journal of Psychology, 71*, 449–474.

Hunt, E. (1983). On the nature of intelligence. *Science, 219*, 141–146.

Hunt, E., Lunneborg, C., & Lewis, J. (1975). What does it mean to be high verbal? *Cognitive Psychology, 7*, 194–227.

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using w2 and d. *American Psychologist, 36*, 892–901.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*, 53–69.

Hyde, J. S., & Linn, M. C. (Eds.) (1986). *The psychology of gender advances through meta-analysis*. Baltimore: Johns Hopkins University Press.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139–155.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39, 1–123.

Jensen, A. R. (1979). G: Outmoded theory or unconquered frontier? *Creative Science Technology, 2*, 16–29.

Jensen, A. R. (1982). Reaction time and psychometric g. In H. J. Eysenck (Eds.), *A model for intelligence* (pp. 93–123). Berlin: Springer-Verlag.

Jensen, A. R. (1993). Why is reaction time correlated with psychometric g? *Current Directions in Psychological Science, 2*, 53–56.

Keating, D. P. (1982). The emperor's new clothes: The "new look" in intelligence research. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 1–45).

Keating, D. P., & Bobbit, B. L. (1978). Individual and developmental differences in cognitive-processing components of mental ability. *Child Development, 49*, 155–167.

Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105*, 198–214.

Kimura, D. (1992). Sex differences in the brain. *Scientific American, 267*, 118–125.

Kimura, D., & Hampson, E. (1994). Cognitive pattern in men and women is influenced by fluctuation in sex hormones. *Current Directions in Psychological Science, 3*, 57–61.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Languages, 30*, 580–602.

Klaus, R. A., & Gray, S. (1968). The early training project for disadvantaged children: A report after five years. *Monographs of the Society for Research in Child Development, 33* (Serial No. 120).

Kline, P. (1991). *Intelligence: The psychometric view*. New York: Routledge.

Kosslyn, S. M., Brunn, J., Cave, K. R., & Wallach, R. W. (1984). Individual differences in mental imagery ability: A computational analysis. *Cognition, 18*, 195–243.

Lachman, R., Lachman, J. L., & Butterfield, E. C. (1979). *Cognitive psychology and information processing: An introduction*. Hillsdale, NJ: Erlbaum.

Larson, G. E., & Saccuzzo, D. P. (1989). Cogntive correlates of general intelligence: Toward a process theory of g. *Intelligence, 13*, 5–31.

Lave, J. (1988). *Cognition in practice*. Cambridge, England: Cambridge University Press.

Law, D. J., Pellegrino, J. W., & Hunt, E. B. (1993). Comparing the tortoise and the hare. Gender differences and experience in dynamic spatial reasoning tasks. *Psychological Science, 41*, 35–40.

Lazar, I., Darlington, R., Murray, H., Royce, J., & Snipper, A. (1982). Lasting effects of early education: A report from the Consortium for Longitudinal Studies. *Monographs of the Society for Research in Child Development, 47* (Serial No. 195).

Loehlin, J. C., Lindzey, G., & Spuhler, J. N. (1975). *Race differences in intelligence*. San Francisco: Freeman.

Longstreth, L. (1984). Jensen's reaction-time investigations of intelligence: A critique. *Intelligence, 8*, 139–160.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

MacKinnon, D. W. (1962). The nature and nurture of creative talent. *American Psychologist, 17*, 484–495.

Mackintosh, J. J. (1986). The biology of intelligence? *British Journal of Psychology, 77*, 1–18.

McClelland, D. C. (1973). Testing for competence rather than for "intelligence." *American Psychologist, 28*, 1–14.

McDermid, C. D. (1965). Some correlates of creativity in engineering personnel. *Journal of Applied Psychology, 49*, 14–19.

Miller-Jones, D. (1989). Culture and testing. *American Psychologist, 44*, 360–366.

Miller, L. B., & Bizzell, R. P. (1984). Long-term effects of four preschool programs: Ninth and tenth-grade results. *Child Development, 55*, 1570–1587.

Mumaw, R. J., Pellegrino, J. W., Kail, R. V., & Carter, P. (1984). Different slopes for different folks: Process analysis of spatial aptitude. *Memory and Cognition, 12*, 515–521.

Neisser, U., Boodoo, G., Bouchard, T. J., Boyken, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbiva, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77–101.

Nettlebeck, T., & Lally, M. (1976). Inspection time and measured intelligence. *British Journal of Psychology, 67*, 17–22.

Newcombe, N., & Bandura, M. M. (1983). Effects of age at puberty on spatial ability in girls: A question of mechanism. *Developmental Psychology, 19*, 215–244.

Oakland, T. (1983). Joint use of adaptive behavior and IQ to predict achievement. *Journal of Consulting and Clinical Psychology, 51*, 298–301.

Ogbu, J. U. (1978). *Minority education and caste: The American system in cross-cultural perspective*. New York: Academic Press.

Ogbu, J. U. (1994). From cultural differences to differences in cultural frames of reference. In P. M. Creenfield & R. R. Cocking (Eds.), *Cross-cultural roots of minority child development* (pp. 365–391). Hillsdale, NJ: Erlbaum.

Parks, R. W., Loewenstein, D. A., Dodril, K. L., Barker, W. W., Yoshi, F., Chang, J. Y., Emran, A., Apicella, A., Shermata, W. A., & Duara, R. (1988). Cerebral metabolic effects of a verbal fluency test: A PET scan study. *Journal of Clinical and Experimental Neuropsychology, 10*, 565–575.

Pellegrino, J. W., & Glaser, R. (1979). Cognitive correlates and components in the analysis of individual differences. *Intelligence, 3*, 187–214.

Perfetto, G. A., & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 141–183). Hillsdale, NJ: Erlbaum.

Perkins, D. N. (1988). Creativity and the quest for mechanism. In R. J. Sternberg & E. E. Smith (Eds.), *The psychology of human thought*. Cambridge, England: Cambridge University Press.

Ramey, C. T., Campbell, F. A., & Finkelstein, N. W. (1984). Course and structure of intellectual development in children at risk for developmental retardation. In P. H. Brooks, R. Sperber, & C. McCauley (Eds.), *Learning and cognition in the mentally retarded*. Hillsdale, NJ: Erlbaum.

Raymond, C. L., & Benbow, C. P. (1986). Gender differences in mathematics: A function of parental support and student sex typing? *Developmental Psychology, 22*, 808–819.

Reed, T. E., & Jensen, A. R. (1992). Conduction velocity in a brain nerve pathway of normal adults correlates with intelligence level. *Intelligence, 16*, 259–272.

Resnick, S. M., Berenbaum, S. A., Gottesman, I. F., & Bouchard, T. J., Jr. (1986). Early hormonal influence on cognitive functioning in congential adrenal hyperplasia. *Developmental Psychology, 22*, 191–198.

Richards, R., Kinney, D. K., Benet, M., & Merzel, A. P. C. (1988). Assessing everyday creativity characteristics of the lifetime creativity scales and validation with three large samples. *Journal of Personality and Social Psychology, 54*, 476–485.

Richardson, K. (1991). *Understanding intelligence*. Philadelphia: Open University Press.

Roe, K. V., Drivas, A., Karagellis, A., & Roe, A. (1985). Sex differences in vocal interaction with mother and stranger in Greek infants: Some cognitive implications. *Developmental Psychology, 21*, 372–377.

Rogoff, B., & Morelli, G. (1989). Perspectives on children's development from cultural psychology. *American Psychologist, 44*, 343–348.

Rothstein, M. G., Paunonen, S. V., Rush, J. C., & King, G. A. (1994). Personality and cognitive ability indicators of performance in graduate business school. *Journal of Educational Psychology, 86*, 516–530.

Ruchalla, E., Scholt, E., & Vogel, F. (1985). Relations between mental performance and reaction time: New aspects of an old problem. *Intelligence, 9*, 189–205.

Rushton, J. P. (1988). Race differences in behaviour: A review and evolutionary analysis. *Personality and Individual Differences, 9*, 1009–1024.

Rushton, J. P. (1991). Do r-K strategies underlie human race differences? *Canadian Psychology, 32*, 29–42.

Scarr, S., Pakstis, A. J., Katz, S. H., & Barker, B. (1977). Absence of a relationship between degree of white ancestry and intellectual skills within a Black population. *Human Genetics, 39*, 69–86.

Scarr, S., & Weinberg, R. A. (1976). IQ test performance of black children adopted by white families. *American Psychologist, 31*, 726–739.

Scarr, S., & Weinberg, R. A. (1983). The Minnesota adoption studies: Genetic differences and malleability. *Child Development, 54*, 260–267.

Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science, 174*, 1285–1295.

Schofield, N. J., & Ashman, A. F. (1986). The relationship between digit span and cognitive processing across ability groups. *Intelligence, 10*, 59–73.

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Constable, R. T., Skudlarski, P., Fulbright, R. K., Bronen, R. A., Fletcher, J. M., Shankweller, D. P., Katz, L., & Gore, J. C. (1995). Sex differences in the functional organization of the brain for language. *Nature, 373*, 607–609.

Shields, J. (1962). *Monoygotic twins brought up apart and brought together*. London: Oxford University Press.

Shuter-Dyson, R. (1982). Musical ability. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic Press.

Silverman, I., & Eals, M. (1992). Sex differences in spatial abilities: Evolutionary theory and data. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 539–549). New York: Oxford University Press.

Smith, C. B. (1984). Aging and changes in cerebral energy metabolism. *Trends in Neurosciences, 7*, 203–208.

Spearman, C. (1927). *The abilities of man*. New York: Macmillan.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge, MA: Cambridge University Press.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence*. Cambridge, England: Cambridge University Press.

Sternberg, R. J., Wagner, R. K., Williams, W. M., & Horvath, J. A. (1995). Testing common sense. *American Psychologist, 50*, 912–927.

Stevenson, H. W., Stigler, J. W., Lee, S., Lucker, G. W., Kitamura, S., & Hsu, C. (1985). Cognitive performance and academic achievement of Japanese, Chinese, and American children. *Child Development, 56*, 718–734.

Tavris, C. (1992). *The mismeasure of woman: Why women are not the better sex, the inferior sex, or the opposite sex*. New York: Simon & Schuster.

Thorndike, E. L. (1936). Factor analysis of social and abstract intelligence. *Journal of Educational Psychology, 27*, 231–233.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press, Psychometric Monographs, No. 1.

Tizard, B., Cooperman, O., Joseph, A., & Tizard, J. (1972). Environmental effects of language development: A study of young children in long-stay residential nurseries. *Child Development, 43*, 337–358.

Van Court, M., & Bean, F. D. (1985). Intelligence and fertility in the United States: 1912–1982. *Intelligence, 9*, 23–32.

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence, 7*, 53–70.

Vernon, P. A., & Kantor, L. (1986). Reaction time correlates with intelligence test scores obtained under either timed and untimed conditions. *Intelligence, 10*, 315–330.

Vernon, P. A., & Mori, M. (1992). Intelligence, reaction times, and peripheral nerve conductor velocity. *Intelligence, 16*, 273–288.

Voyer, D., Voyer, S., & Bryden, M. F. (1995). Magnitude of sex difference in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*, 250–270.

Waber, D. P., Mann, M. B., Merola, J., & Moylan, P. (1985). Physical maturation rate and cognitive performance in early adolescence: A longitudinal examination. *Developmental Psychology, 21*, 666–681.

Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology, 49*, 436–458.

Wagner, R. K., & Sternberg, R. J. (1990). Streetsmarts. In K. E. Clark & M. B. Clark (Eds.), *Measures of leadership* (pp. 493–504). West Orange, NJ: Leadership Library of American.

Wallach, M. A. (1976). Tests tell us little about talent. *American Scientist, 64*, 57–63.

White, B. L. (1978). *Experience and environment: Major influences on the development of the young child* (Vol. 2). Englewood Cliffs, NJ: Prentice Hall.

Wissler, C. (1901). The correlation of mental and physical tests. *Psychological Review Monograph Supplements, 3(6)*, Whole No. 16.

Wong, G. M. T., Day, J. D., Maxwell, S. E., & Meara, N. M. (1995). A multitrait-multimethod study of academic and social intelligence in college students. *Journal of Educational Psychology, 87*, 117–133.

Woodrow, H. (1939). The common factos in fifty-two mental tests. *Psychometrika, 4*, 99–108.

Yekovich, F. R., Walker, C. H., Ogle, L. T., & Thompson, M. A. (1990). The influence of domain knowledge on inferencing in low-aptitude individuals. *The Psychology of Learning and Motivation, 25*, 259–278.

Yong, L. M. S. (1994). Relations between creativity and intelligence among Malaysian pupils. *Perceptual and Motor Skills, 79*, 739–742.

PART XVIII

Cognitive Neuroscience

Chapter 37

Localization of Cognitive Operations in the Human Brain

Michael I. Posner, Steven E. Petersen, Peter T. Fox, and Marcus E. Raichle

Introduction

The question of localization of cognition in the human brain is an old and difficult one (Churchland, 1986). However, current analyses of the operations involved in cognition (Anderson, 1980) and new techniques for the imaging of brain function during cognitive tasks (Raichle, 1983) have combined to provide support for a new hypothesis. The hypothesis is that elementary operations forming the basis of cognitive analyses of human tasks are strictly localized. Many such local operations are involved in any cognitive task. A set of distributed brain areas must be orchestrated in the performance of even simple cognitive tasks. The task itself is not performed by any single area of the brain, but the operations that underlie the performance are strictly localized. This idea fits generally with many network theories in neuroscience and cognition. However, most neuroscience network theories of higher processes (Mesolam, 1981; Goldman-Rakic, 1988) provide little information on the specific computations performed at the nodes of the network, and most cognitive network models provide little or no information on the anatomy involved (McClelland & Rumelhart, 1986). Our approach relates specific mental operations as developed from cognitive models to neural anatomical areas.

The study of reading and listening has been one of the most active areas in cognitive science for the study of internal codes involved in information processing (Posner, 1986). In this chapter we review results of studies on cognitive tasks that suggest several separate codes for processing individual words. These codes can be accessed from input or from attention. We also review studies of alert monkeys and brain-lesioned patients that provide evidence on the localization of an attention system for visual spatial information. This system is apparently unnecessary for processing single, foveally centered words. Next, we introduce data from positron emission tomography (PET) concerning the neural systems underlying the coding of individual visual (printed) words. These studies support the findings in cognition and also give new evidence for an anterior attention system involved in language processing. Finally, we survey other areas of cognition for which recent findings support the localization of component mental operations.

From *Science* 240 (1988): 1627–1631. Reprinted with permission.

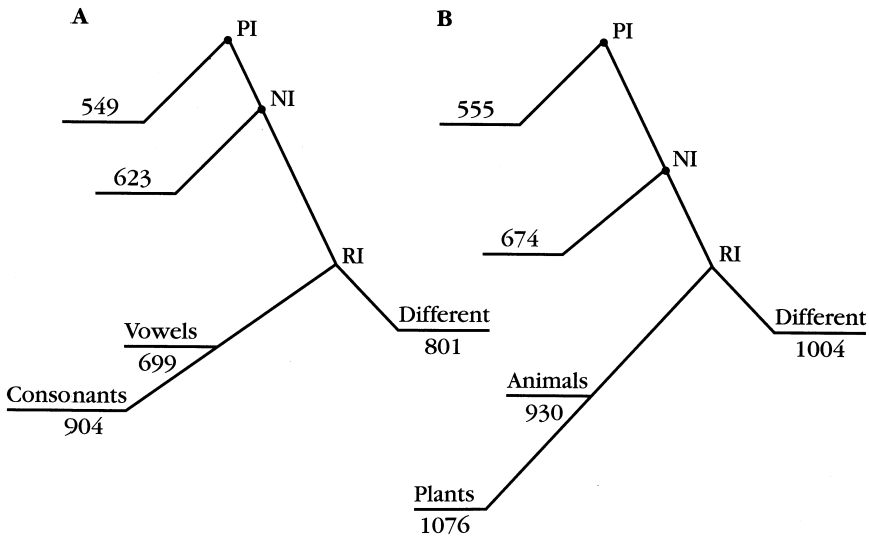


Figure 37.1
Results of reaction time studies in which subjects were asked to classify whether pairs of letters were both vowels or both consonants (A) or whether pairs of words were both animal or both plants (B). Reaction times are in milliseconds. Each study involved 10 to 12 normal subjects. Standard deviations are typically 20% of the mean value. Data argue in favor of these matches being made on different internal codes (Posner, 1986; Marshall & Newcombe, 1973; LaBerge & Samuels, 1974; Carr & Pollatsek, 1985; Coltheart, 1985). Abbreviations: PI, physical identity; NI, name identity; and RI, rule identity. [Reprinted from Posner, Lewis, & Conrad (1972) with permission of MIT Press.]

Internal Codes

The most advanced efforts to develop cognitive models of information processing have been in the area of the coding of individual words through reading and listening (Posner, 1986; Marshall & Newcombe, 1973; LaBerge & Samuels, 1974; Carr & Pollatsek, 1985; Coltheart, 1985). These efforts have distinguished between a number of internal codes related to the visual, phonological, articulatory, and semantic analysis of a word. Operations at all these levels appear to be involved in understanding a word.

This view began with efforts to develop detailed measurements of the time it takes to execute operations on codes thought to be involved in reading. Figure 37.1 shows the amount of time needed to determine if two simultaneously shown visual letters or words belong to the same category (Posner, Lewis, & Conrad, 1972). The reaction time to match pairs of items that are physically identical (for example, AA) is faster than reaction time for matches of the same letters or words in the opposite case (Aa), which are in turn faster than matches that have only a common category (Ae). These studies have been interpreted as involving a mental operation of matching based on different codes. In the case of visual identity the code is thought to be the visual form, whereas in cross-case matching it is thought to be the letter or word name. The idea that a word consists of separable physical, phonological, and semantic codes and that operations may be performed on them separately has been basic to many

theories of reading and listening (Marshall & Newcombe, 1973; LaBerge & Samuels, 1974; Carr & Pollatsek, 1985; Coltheart, 1985). Thus the operation of rotating a letter to the upright position is thought to be performed on the visual code (Cooper, 1976), whereas matching to determine if two words rhyme is said to be performed on a phonological representation of the words (Kleiman, 1975). These theories suggest that mental operations take place on the basis of codes related to separate neural systems.

It is not easy to determine if any operation is elementary or whether it is based on only a single code. Even a simple task such as matching identical items can involve parallel operations on both physical and name codes. Indeed, there has been controversy over the theoretical implications of these matching experiments (Boles & Everland, 1981). Some results have suggested that both within- and cross-case matches are performed on physical (visual) codes, whereas others have suggested that they are both performed on name codes (Boles & Everland, 1981). A basic question is to determine whether operations performed on different codes involve different brain areas. This question cannot be resolved by performance studies, since they provide only indirect evidence about localization of the operations performed on different codes.

It has been widely accepted that there can be multiple routes by which codes interact. For example, a visual word may be sounded out to produce a phonological code and then the phonology is used to develop a meaning (Posner, 1986; Marshall & Newcombe, 1973; LaBerge & Samuels, 1974; Carr & Pollatsek, 1985; Coltheart, 1985). Alternately, the visual code may have direct access to a semantic interpretation without any need for developing a phonological code (Posner, 1986; Marshall & Newcombe, 1973; LaBerge & Samuels, 1974; Carr & Pollatsek, 1985; Coltheart, 1985). These routes are thought to be somewhat separate because patients with one form of reading difficulty have great trouble in sounding out nonsense material (for example, the nonword "caik"), indicating they may have a poor ability to use phonics; but they have no problems with familiar words even when the words have irregular pronunciation (for example, pint). Other patients have no trouble with reading nonwords but have difficulty with highly familiar irregular words. Although there is also reason to doubt that these routes are entirely separate, it is often thought that the visual to semantic route is dominant in skilled readers (Marshall & Newcombe, 1973; LaBerge & Samuels, 1974; Carr & Pollatsek, 1985; Coltheart, 1985).

Visual Spatial Attention

Another distinction in cognitive psychology is between automatic activation of these codes and controlled processing by means of attention (Posner, 1986; Marshall & Newcombe, 1973; LaBerge & Samuels, 1974; Carr & Pollatsek, 1985; Coltheart, 1985). Evidence indicates that a word may activate its internal visual, phonological, and even semantic codes without the person having to pay attention to the word. The evidence for activation of the internally stored visual code of a word is particularly good. Normal subjects show evidence that the stimulus duration necessary for perceiving individual letters within words is shorter than for perceiving the same letter when it is presented in isolation (Reicher, 1969; McClelland & Rumelhart, 1981).

What is known about the localization of attention? Cognitive, brain lesion, and animal studies have identified a posterior neural system involved in visual spatial attention. Patients with lesions of many areas of the brain show neglect of stimuli from the side of space opposite the lesion (DeRenzi, 1982). These findings have led to network views of the neural system underlying visual spatial attention (Mesulam, 1984). However, studies performed with single-cell recording from alert monkeys have been more specific in showing three brain areas in which individual cells show selective enhancement due to the requirement that the monkey attend to a visual location (Mountcastle, 1978; Wurtz, Goldberg, & Robinson, 1980; Petersen & Robinson, 1985). These areas are the posterior parietal lobe of the cerebral cortex, a portion of the thalamus (part of the pulvinar), and areas of the midbrain related to eye movements—all areas in which clinical studies of lesioned patients find neglect of the environment opposite the lesion.

Recent studies of normal (control) and patient populations have used cues to direct attention covertly to areas of the visual field without eye movements (Posner, Walker, Friedrich, & Rafal, 1984). Attention is measured by changes in the efficiency of processing targets at the cued location in comparison with other uncued locations in the visual field. These studies have found systematic deficits in shifting of covert visual attention in patients with injury of the same three brain areas suggested by the monkey studies. When the efficiency of processing is measured precisely by a reaction time test, the nature of the deficits in the three areas differs. Patients with lesions in the parietal lobe show very long reaction times to targets on the side opposite the lesion only when their attention has first been drawn to a different location in the direction of the lesion (Posner, Walker, Friedrich, & Rafal, 1984). This increase in reaction time for uncued but not cued contralesional targets is consistent with a specific deficit in the patient's ability to disengage attention from a cued location when the target is in the contralesional direction. In contrast, damage to the midbrain not only greatly lengthens overall reaction time but increases the time needed to establish an advantage in reaction time at the cued location in comparison to the uncued location (Posner, Cohen, & Rafal, 1982). This finding in consistent with the idea that the lesion causes a slowing of attention movements. Damage to the thalamus (Rafal & Posner, 1987) produces a pattern of slowed reaction to both cued and uncued targets on the side opposite the lesion. This pattern suggests difficulty in being able to use attention to speed processing of targets irrespective of the time allowed to do so (engage deficit). A similar deficit has been found in monkeys performing this task when chemical injections disrupt the performance of the lateral pulvinar (Petersen, Robinson, & Reys, 1985). Thus the simple act of shifting attention to the cued location appears to involve a number of distinct computations (figure 37.2) that must be orchestrated to allow the cognitive performance to occur. We now have an idea of the anatomy of several of these computations.

Damage to the visual spatial attention system also produces deficits in recognition of visual stimuli. Patients with lesions of the right parietal lobe frequently neglect (fail to report) the first few letters of a nonword. However, when shown an actual word that occupied the same visual angle, they report it correctly (Sieroff, Palatsek, & Posner, 1988). Cognitive studies have often

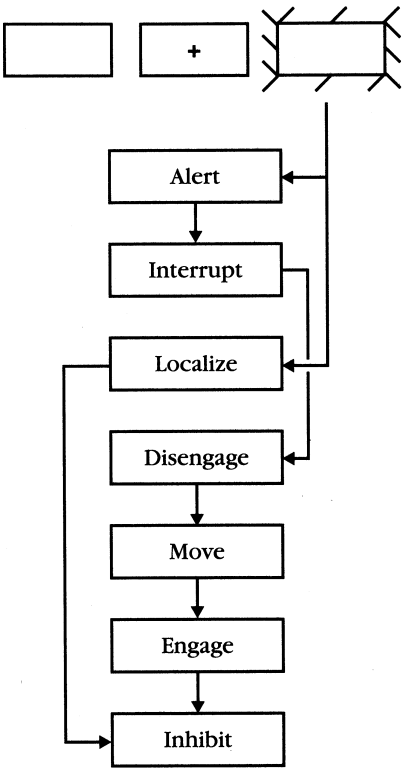


Figure 37.2
Top of figure illustrates an experimental situation in which attention is summoned from fixation (center) to righthand box by brightening of the box. This is followed by a target at the cued location or on the opposite side. The boxes below indicate mental operations thought to begin by presentation of the cue. The last four operations involve the posterior visual–spatial attention system; specific deficits have been found in patients with lesions in the parietal (disengage), midbrain (move), and thalmic (engage) areas (Posner, Walker, Friedrich, & Rafal, 1984; Posner, Cohen, & Rafal, 1982; Rafar and Posner, 1987). [Reprinted from Posner, Inhoff, Friedrich, & Cohen (1987) with permission of the Psychonomics Society.]

shown a superiority of words over nonwords (Reicher, 1969; McClelland & Rumelhart, 1984). Our results fit with the idea that words do not require scanning by a covert visual spatial attention system.

Attention for Action

In cognitive studies it is often suggested that attention to stimuli occurs only after they have been processed to a very high degree (Allport, 1980; Duncan, 1980). In this view, attention is designed mainly to limit the conflicting actions taken toward stimuli. This form of attention is often called "attention for action." Our studies of patients with parietal lesions suggest that the posterior visual spatial attention system is connected to a more general attention system that is also involved in the processing of language stimuli (Posner, Inhoff, Friedrich, & Cohen, 1987). When normal subjects and patients had to pay close

Table 37.1
Conditions for PET subtractive studies of words

Control state	Stimulus state	Computations
Fixation	Passive words	Passive word processing
Repeat words	Generate word use	Semantic association, attention
Passive words	Monitor category	Semantic association, attention (many targets)[a]

[a] The extent of attentional activation increases with the number of targets.

attention to auditory, or spoken, words, the ability of a visual cue to draw their visual spatial attention was retarded. Cognitive studies have been unclear on whether access to meaning requires attention. Although semantic information may be activated without attention being drawn to the specific lexical unit (Marcel, 1983), attention strongly interacts with semantic activation (Henik, Friedrich, & Kellogg, 1983; Hoffman & Macmillan, 1985). Considerable evidence shows that attention to semantic information limits the range of concepts activated. When a person attends to one meaning of a word, activation of alternative meanings of the same item tends to be suppressed (Neely, 1977).

PET Imaging of Words

How do the operations suggested by cognitive theories of lexical access relate to brain systems? Recently, in a study with normal persons, we used PET to observe brain processes that are active during single word reading (Petersen, Fox, Posner, Mintun, & Raichle, 1988). This method allows examination of averaged changes in cerebral blood flow in localized brain areas during 40 seconds of cognitive activity (Fox, Mintun, Reiman, & Raichle, 1988). During this period we presented words at a rate of one per second. Previous PET studies have suggested that a difference of a few millimeters in the location of activations will be sufficient to separate them (Fox et al., 1986).

To isolate component mental operations we used a set of conditions shown in table 37.1. By subtracting the control state from the stimulus state, we attempted to isolate areas of activation related to those mental operations present in the stimulus state but not in the control state. For example, subtraction of looking at the fixation point, without any stimuli, from the presentation of passive visual words allowed us to examine the brain areas automatically activated by the word stimuli.[1]

Visual Word Forms

We examined changes in cerebral blood flow during passive looking at foveally presented nouns. This task produced five areas of significantly greater activation than found in the fixation condition. They all lie within the occipital lobe: two along the calcarine fissure in left and right primary visual cortex and three in left and right lateral regions (figure 37.3). As one moves to more complex naming and semantic activation tasks, no new posterior areas are active. Thus the entire visually specific coding takes place within the occipital lobe. Activated areas are found as far anterior as the occipital temporal boundary. Are these activations specific to visual words? The presentation of auditory words

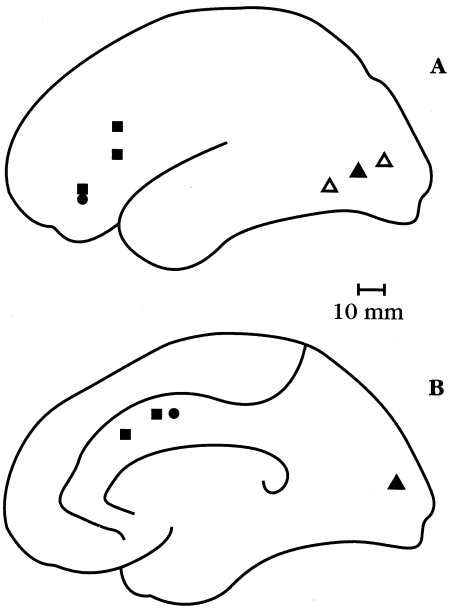


Figure 37.3
Areas activated in visual word reading on the lateral aspect of the cortex (A) and on the medical aspect (B). Triangles refer to the passive visual task minus fixation (black triangle, left hemisphere; white triangle, right hemisphere). Only occipital areas are active. Squares refer to generate minus repeat task. Circles refer to monitor minus passive words task. Solid circles and squares in (A) denote left hemisphere activation; however, in (B), on the midline it is not possible to determine if activation is left or right. The lateral area is thought to involve a semantic network while the midline areas appear to involve attention (Petersen, Fox, Posner, Mintun, & Raichle, 1988).

does not produce any activation in this area. Visual stimuli known to activate striate cortex (for example, checkerboards or dot patterns) do not activate the prestriate areas used in word reading (Fox et al., 1986; Fox, Miezin, Allman,Van Essen, & Raichle, 1987). All other cortical areas active during word reading are anterior. Thus it seems reasonable to conclude that visual word forms are developed in the occipital lobe.

It might seem that occipital areas are too early in the system to support the development of visual word forms. However, the early development of the visual word form is supported by our evidence that patients with right parietal lesions do not neglect the left side of foveally centered words even though they do neglect the initial letters of nonword strings (Sieroff, Pollatsek, & Posner, 1988). The presence of pure alexia from lesions of the occipital temporal boundary (Damasio & Damasio, 1983) also supports the development of the visual word form in the occipital area.

Precise computational models of how visual word forms are developed (McClelland & Rumelhart, 1986; Reicher, 1969; McClelland & Rumelhart, 1981) involve parallel computations from feature, letter, and word levels and precise feedback among these levels. The prestriate visual system would provide an attractive anatomy for models relying on such abundant feedback. However,

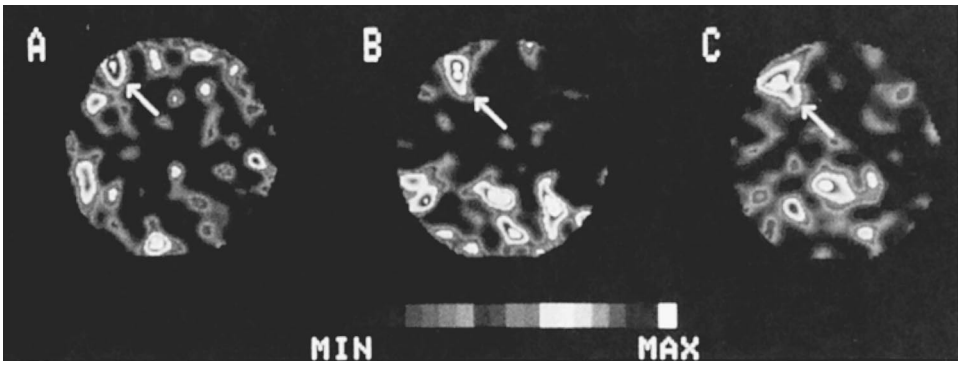


Figure 37.4
Sample data from the PET activation studies. The arrows indicate areas of activation in the left inferior prefrontal cortex found active in all three semantic processing conditions. (Left) Monitoring visual words for dangerous animals (minus passive visual words). (Middle) Generating uses (minus repeat) for visual stimuli. (Right) Generating uses (minus repeat) for auditory stimuli. In each condition an area of cortical activation was found in the anterior cingulate gyrus on a higher slice (figure 37.3). The color scale indicates the relative strength of activation (black indicates the minimum and white, the maximum, for that condition; Petersen, Fox, Posner, Mintun, & Raichle, 1988).

presently we can only tentatively identify the general occipital areas that underlie the visual processing of words.

Semantic Operations

We used two tasks to study semantic operations. One task required the subject to generate and say aloud a use for each of 40 concrete nouns (for example, a subject may say "pound" when presented with the noun "hammer"). We subtracted the activations from repeating the nouns to eliminate strictly sensory and motor activations. Only two general areas of the cortex were found to be active (figure 37.3, square symbols). A second semantic task required subjects to note the presence of dangerous animals in a list of 40 visually presented words. We subtracted passive presentation of the word list to eliminate sensory processing. No motor output was required and subjects were asked to estimate only the frequency of targets after the list was presented. The same two areas of cortex were activated (figure 37.3, circles).

One of the areas activated in both semantic tasks was in the anterior left frontal lobe. Figure 37.4 shows an illustration of this area from averaged scans in auditory and visual generate (minus repeat) and in visual monitoring (minus passive words). This area is strictly left lateralized and appears to be specific to semantic language tasks. Moreover, lesions of this area produce deficits in word fluency tests (Benton, 1968). Thus we have concluded that this general area is related to the semantic network supporting the type of word associations involved in the generate and monitoring tasks.

Phonological Coding

When words are presented in auditory form, the primary auditory cortex and an area of the left temporoparietal cortex that has been related to language

tasks are activated (Geschwind, 1965). This temporoparietal left-lateralized area seemed to be a good candidate for phonological processing. It was surprising from some perspectives that no visual word reading task activated this area. However, all of our visual tasks involved single common nouns read by highly skilled readers. According to cognitive theories of reading (Marshall & Newcombe, 1973; LaBerge & Samuels, 1974; Carr & Pollatsek, 1985; Coltheart, 1985), these tasks should involve the visual to semantic route. One way of requiring a phonological activation would be to force subjects to tell whether two simultaneous words (for example, pint-lint or row-though) rhymed. This method has been used in cognitive studies to activate phonological codes (Kleiman, 1975). Recent data from our laboratory (Petersen, Fox, Posner, & Raichle, unpub.) show that this task does produce activation near the supramarginal gyrus. We also assume that word reading that involves difficult words or requires storage in short-term memory or is performed by unskilled readers would also activate phonological operations.

Anterior Attention

There is no evidence of activation of any parts of the posterior visual spatial attention system (for example, parietal lobe) in any of our PET language studies. However, it is possible to show that simple tasks that require close monitoring of visual input or that use visual imagery (Petersen, Fox, Miezin, & Raichle, 1988) do activate this parietal system. We conclude, in agreement with the results of our lesion work (Sieroff, Pollatsek, & Posner, 1988), that visual word reading is automatic in that it does not require activation of the visual spatial attention system.

In recent cognitive theories the term *attention for action* is used to summarize the idea that attention seems to be involved in selecting those operations that will gain control of output systems (Allport, 1980). This kind of attention system does not appear to be related to any particular sensory or cognitive content and is distinct from the more strictly visual functions assigned to the visual-spatial attention system. Although attention for action seems to imply motor acts, internal selections involved in detecting or noting an event may be sufficient to involve attention in this sense (Duncan, 1980). Whenever subjects are active in this way, we see an increase in blood flow in areas of the medial frontal lobe (figure 37.3B, square symbols). When motor output is involved (for example, naming words), these areas tend to be more superior and posterior (supplementary motor area); but when motor activity is subtracted away or when none is required, they appear to be more anterior and inferior (anterior cingulate gyrus). The anterior cingulate has long been thought to be related to attention (Mesulam, 1988) in the sense of generating actions, since lesions of this area produce akinetic mutism (Damasio & Van Hoesen, 1983).

We tested the identification of the anterior cingulate with attention and the left lateral frontal area with a word association network. This was done by applying a cognitive theory that attention would not be much involved in the semantic decision of whether a word belonged to a category (for example, dangerous animal) but would be involved in noting the targets even though no specific action was required. The special involvement of attention with target detection has been widely argued by cognitive studies (Duncan, 1980). These

studies have suggested that monitoring produces relatively little evidence of heavy attentional involvement, but when a target is actually detected there is evidence of strong interference so that the likelihood of detecting a simultaneous target is reduced. Thus we varied the number of dangerous animals in our list from one (few targets) to 25 (many targets). We found that blood flow in the anterior cingulate showed much greater change with many targets than with few targets. The left frontal area showed little change in blood flow between these conditions. Additional work with other low-target vigilance tasks not involving semantics also failed to activate the anterior cingulate area.[2] Thus the identification of the anterior cingulate with some part of an anterior attention system that selects for action receives some support from these results.

Conclusions

The PET data provide strong support for localization of operations performed on visual, phonological, and semantic codes. The ability to localize these operations in studies of average blood flow suggests considerable homogeneity in the neural systems involved, at least among the right-handed subjects with good reading skills who were used in our study.

The PET data on lexical access complement the lesion data cited here in showing that mental operations of the type that form the basis of cognitive analysis are localized in the human brain. This form of localization of function differs from the idea that cognitive tasks are performed by a particular brain area. Visual imagery, word reading, and even shifting visual attention from one location to another are not performed by any single brain area. Each of them involves a large number of component computations that must be orchestrated to perform the cognitive task.

Our data suggest that operations involved both in activation of internal codes and in selective attention obey the general rule of localization of component operations. However, selective attention appears to use neural systems separate from those involved in passively collecting information about a stimulus. In the posterior part of the brain, the ventral occipital lobe appears to develop the visual word form. If active selection or visual search is required, this is done by a spatial system that is deficient in patients with lesions of the parietal lobe (Friedrich, Walker, & Posner, 1985; Riddoch & Humphreys, 1987). Similarly, in the anterior brain the lateral left frontal lobe is involved in the semantic network for coding word associations. Local areas within the anterior cingulate become increasingly involved when the output of the computations within the semantic network is to be selected as a relevant target. Thus the anterior cingulate is involved in the computations in selecting language or other forms of information for action. This separation of anterior and posterior attention systems helps clarify how attention can be involved both in early visual processing and in the selection of information for ontput.

Several other research areas also support our general hypothesis. In the study of visual imagery, models distinguish between a set of operations involved in the generation of an image and those involved in scanning the image once it is generated (Kosslyn, 1980). Mechanisms involved in image scanning share

components with those in visual spatial attention. Patients with lesions of the right parietal lobe have deficits both in scanning the left side of an image and in responding to visual input to their left (Bisiach & Luzzatti, 1978). Although the right hemisphere plays an important role in visual scanning, it apparently is deficient in operations needed to generate an image. Studies of patients whose cerebral hemispheres have been split during surgery show that the isolated left hemisphere can generate complex visual images whereas the isolated right hemisphere cannot (Kosslyn, Holtzman, Farah, & Gazzaniga, 1985).

Patients with lesions of the lateral cerebellum have a deficit in timing motor output and in their threshold for recognition of small temporal differences in sensory input (Ivry, Keele, & Diener, 1988). These results indicate that this area of the cerebellum performs a critical computation for timing both motor and sensory tasks. Similarly, studies of memory have indicated that the hippocampus performs a computation needed for storage in a manner that will allow conscious retrieval of the item once it has left current attention. The same item can be used as part of a skill even though damage to the hippocampus makes it unavailable to conscious recollection (Squire, 1986).

The joint anatomical and cognitive approach discussed in this article should open the way to a more detailed understanding of the deficits found in the many disorders involving cognitive or attentional operations in which the anatomy is poorly understood. For example, we have attempted to apply the new knowledge of the anatomy of selective attention to study deficits in patients with schizophrenia (Early, Posner, & Reiman, Posner, Early, Reiman, Pardo, et al., 1988).

Notes

This work was supported by the Office of Naval Research contract N00014-86-K-0289 and by the McDonnell Center for Higher Brain Studies. The imaging studies were performed at the Malinckrodt Institute of Radiology of Washington University with the support of NIH grants NS 06833, HL 13851, NS 14834, and AG 03991. We thank M. K. Rothbart and G. L. Shulman for helpful comments.

1. Subtraction was used to infer mental processes by F. C. Donders in 1868 for reaction time data. The method has been disputed because it is possible that subjects use different strategies as the task is made more complex. By using PET, we can study this issue. For example, when subtracting the fixation control from the generate condition, one should obtain only those active areas found in passive (minus fixation) plus repeat (minus passive) plus generate (minus repeat). Our preliminary analyses of these conditions generally support the method.
2. The studies of the visual monitoring task were conducted by S. E. Petersen, P. T. Fox, M. I. Posner, and M. E. Raichle. Unpublished studies on vigilance were conducted by J. Pardo, P. T. Fox, M. I. Posner, and M. E. Raichle, using somatosensory and visual tasks.

References

Allport, D. A. (1980). In G. Claxton (Ed.), *Cognitive psychology: New directions* (pp. 112–153). Boston: Routledge & Kegan Paul.

Anderson, J. R. (1980). *Cognitive psychology and its implications*. San Francisco: Freeman.

Benton, A. L. (1968). *Neuropsychologia 18*, 53.

Bisiach, E., & Luzzatti, C. (1978). *Cortex 14*, 129.

Boles, D. B., & Eveland, D. C. (1983). *J. Exp. Psychol. Hum. Percept. Perform. 9*, 657; Proctor, R. W. (1981). *Psychol. Rev. 88*, 291.

Churchland, P. S. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.

Cooper, L. A. (1976). *Percept. Psychophys*. 7, 20.

Damasio, A. R., & Damasio, H. (1983). *Neurology 33*, 1573.
Damasio, A. R., & Van Hoesen, G. W. (1983). In K. M. Heilman and P. Satz (Eds.), *Neuropsychology of Human Emotion* (pp. 85–110). NewYork: Guilford.
DeRenzi, E. (1982). *Disorders of space exploration and cognition*. NewYork: Wiley.
Duncan, J. (1980). *Psychol. Rev. 87*, 272.
Fox, P. T. et al. (1986). *Nature 323*, 806.
Fox, P. T., Miezin, F. M., Allman, J. M., Van Essen, D. C., & Raichle, M. E. (1987). *J. Neurosci*. 7, 913.
Fox, P. T., Mintun, M. A., Reiman, E. M., Raichle, M. E. (1988). Enhanced detection of focal brain responses using intersubject averaging and change-distribution analysis of subtracted PET images. *J. Cereb. Blood Flow Metab. 8(5)*, 642–653.
Friedrich, F. J., Walker, J. A., & Posner, M. I. (1985). *Cog. Neuropsychol. 2*, 250; Riddoch, J. M., & Humphreys, G. W. (1980). In M. Jeannerod (Ed.), *Neurophysiological and neuropsychological aspects of spatial neglect* (pp. 151–181). New York: Elsevier.
Geschwind, N. (1965). *Brain 88*, 227.
Henik, A., Friedrich, F. J., & Kellogg, W. A. (1983). *Mem. Cognit. 11*, 363; Hoffman, J. E., & Macmillan, F. W. In M. I. Posner and O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 585–599). Hillsdale, NJ: Erlbaum.
Ivry, R. I., Keele, S. W., & Diener, H. C. (1988). *Exp. Brain Res. 73*, 167–180.
Kleiman, G. M. (1975). *J. Verb. Learn. Verb. Behav. 24*, 323.
Kosslyn, S. W. (1980). *Image and mind*. Cambridge, MA: Harvard Univ. Press.
Kosslyn, S. W., Holtzman, J. D., Farah, M. J., & Gazzaniga, M. S. (1985). *J. Exp. Psychol. Gen. 114*, 311.
Marcel, A. (1983). *Cog. Psychol. 15*, 197.
Marshall, J. C., & Newcombe, F. J. (1973). *J. Psychol. Res. 2*, 175; LaBerge, D., & Samuels, J. (1974). *Cog. Psychol. 10*, 293; Carr, T. H., & Pollatsek, A. (1985). In D. Besner, D. Waller, & G. E. Mackinnon (Eds.), *Reading research*, vol. 5. NewYork: Academic Press, pp. 1–82; Coltheart, M. (1985). In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI*, Hillsdale, NJ: Erlbaum, pp. 3–37.
McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing*, vol. 2. Cambridge, MA: MIT Press, pp. 170–215.
Mesulam, M. M. (1981). *Ann. Neurol. 10*, 309; Goldman-Rakic, P. S. (1988). *Annu. Rev. Neurosci. 11*, 156.
Mountcastle, V. B. (1978). *J. R. Soc. Med. 71*, 14; Wurtz, R. H., Goldberg, M. E., & Robinson, D. L. (1980). *Prog. Psychobiol. Physiol. Psychol. 9*, 43; Petersen, S. E., Robinson, D. L., & Keys, W. (1985). *J. Neurophysiol. 54*, 367.
Neely, J. (1977). *J. Exp. Psychol. Gen. 3*, 226.
Petersen, S. E., Fox, P. T., Miezin, F. M., & Raichle, M. E. (1988). *Invest. Ophthalmol. Vis. Sci. 29*, 22 (abstr.).
Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). *Nature 331*, 585.
Petersen, S. E., Fox, P. T., Posner, M. I., & Raichle, M. E., unpublished data.
Petersen, S. E., Robinson, D. L., & Keys, W. J. (1985). *Neurophysiology 54*, 207.
Posner, M. I. (1986). *Chronometric explorations of mind*. Oxford: Oxford Univ. Press.
Posner, M. I., Cohen, Y., & Rafal, R. D. (1982). *Proc. R. Soc. London Ser. B 298*, 187; Posner, M. L., Choate, L., Rafal, R. D., & Vaughan, J. (1985). *Cog. Neuropsychol*. 2, 250.
Posner, M. I., Early, T. S. Reiman, E., Pardo, P., et al. (1988). Asymetries in hemispheric control of attention in schizophrenia. *Archives of General Psychiatry, 45(9)*, 814–821.
Posner, M. I., Inhoff, W. R., Friedrich, F. J., & Cohen, A. (1987). *Psychobiology 15*, 107.
Posner, M. I., Lewis, J., & Conrad, C. (1972). In J. F. Kavanaugh & I. G. Mattingly (Eds.), *Language by ear and by eye*. Cambridge, MA: MIT Press, pp. 159–192.
Posner, M. I., Walker, J., Friedrich, F. J., & Rafal, R. D. (1984). *J. Neurosci. 4*, 1863.
Rafal, R. D., & Posner, M. I. (1987). *Proc. Natl. Acad. Sci. U.S.A. 84*, 7349.
Raichle, M. E. (1983). *Annu. Rev. Neurosci. 6*, 243.
Reicher, G. M. (1969). *J. Exp. Psychol. 81*, 274; McClelland, J. L., & Rumelhart, D. E. (1981). *Psychol. Rev. 88*, 375.
Sieroff, E., Pollastek, A., and Posner, M. (1988) Recognition of visual letter strings following injury to the posterior visual spatial attention system. *Cog. Neuropsychol. 5(4)*, 451–472.
Squire, L. R. (1986). *Science 232*, 1612.

Chapter 38

The Mind and Donald O. Hebb

Peter M. Milner

Donald O. Hebb, one of the most influential psychologists of his time, began his adult life intending to be a novelist. Deciding that his calling required an understanding of psychology, he embarked on a course that led him into two decades of research. His studies culminated in 1949 with the publication of *The Organization of Behavior*, a keystone of modern neuroscience.

The monograph broke new ground by positing neural structures, called cell assemblies, which were formed through the action of what is now called the Hebb synapse. The cell-assembly theory guided Hebb's landmark experiments on the influence of early environment on adult intelligence. It foreshadowed neural network theory, an active line of research in artificial intelligence.

Hebb's book came at the right time because it flew in the face of behaviorism just as that school was losing its dominance. The behaviorists denounced explanations of behavior by association of ideas (which they called mentalism) and by the action of neurons (which they called physiologizing). But many psychologists had grown weary of the artificial theories these strictures had engendered, and they were captivated by Hebb's project and his engaging literary style. The book became a classic, and Hebb became a household word (at least in psychologists' households).

Hebb never claimed that his 1949 theory was firmly grounded in physiology. His model gave workers something to look for, and later, as knowledge of the brain grew, it became possible to frame his ideas in more realistic neural terms. None of this subsequent research has invalidated Hebb's basic hypothesis. Indeed, its influence appears in many areas of current research.

Hebb was born in Chester, a small fishing and boat-building town in Nova Scotia. His parents were physicians, and his two brothers and his sister followed in their parents' footsteps. But Donald demonstrated his independence early by studying English in preparation for a career as a writer, graduating in 1925 from Dalhousie University in Halifax. To earn his living while gestating his first novel, he taught school in his hometown. A year later he set out to see life, going west to work an eight-horse team on prairie farms. Then, failing to get a job as a deckhand on a freighter to China, he returned east and got a job as a laborer in Quebec.

In 1927 an aspiring novelist not only had to know life but also the works of Sigmund Freud. This was Hebb's introduction to psychology. He was sufficiently intrigued to apply to the psychology department of McGill University, where he was accepted in 1928 as a part-time graduate student. Again he

From *Scientific American* (January 1993): 124–129. Reprinted with permission.

supported himself by teaching and, again, what started out as a temporary interest verged on becoming a career. After one year he was made principal of an elementary school in a working-class district of Montreal. He was determined to make learning enjoyable, taking care to prevent schoolwork from being used as a punishment, instead sending miscreants out of class to play in the school yard. Hebb became absorbed in his educational experiments and seriously considered remaining in the profession. Two developments dissuaded him. He came down with a tubercular hip that confined him to bed for a year and left him with a slight limp. Then his bride of 18 months was killed in an automobile accident. He therefore decided to leave Montreal.

While confined to bed, Hebb wrote a master's thesis that involved him in the nature-nurture controversy. The thesis attempted to explain spinal reflexes as the result of Pavlovian conditioning in the fetus. He subsequently buried all references to this essay both because he changed his mind about its content and because he came to oppose psychological research that lacked an experimental foundation.

One of his examiners was Boris P. Babkin, a physiologist who had worked with Pavlov in St. Petersburg. He recommended that Hebb get some experience in the laboratory and arranged for him to work with another Russian emigre, Leonid Andreyev. Hebb conditioned dogs and became less impressed with Pavlovian techniques. After much soul-searching as to whether he should continue in psychology, he decided in 1934 to burn his boats, borrow money and go to Chicago to continue his doctoral research under Karl S. Lashley.

The elder scientist was to exert a profound influence on Hebb's approach, above all in his emphasis on physiology. Lashley had never doubted that to understand behavior one must first understand the brain. As a lab boy in 1910, he had salvaged slides of a frog brain from the trash heap and tried to find in the neural connections some clue to frog behavior. Lashley performed experiments to detect memory traces in the brain, inventing techniques for making brain lesions and measuring their location and extent. By around 1930 he had become convinced that memories could not be stored in a single region of the brain but must be spread throughout. In 1934, when Hebb went to Chicago, Lashley was concentrating on the study of vision.

A year later Lashley was offered a professorship at Harvard University and managed to take Hebb along. Hebb had to start his research from scratch, and having only enough money for one more year, he sought an experiment that could support a thesis no matter how it came out. He contrived to adapt his interest in the nature-nurture question to Lashley's vision project by investigating the effects of early experience on the development of vision in the rat.

Contrary to the empiricist ideas of his master's thesis, Hebb found that rats reared in complete darkness could distinguish the size and brightness of patterns as accurately as rats reared normally. This finding indicated that the organization of the visual system was innate and independent of environmental cues, a view coinciding with that of the Gestalt school, to which Lashley was sympathetic [see "The Legacy of Gestalt Psychology," by Irvin Rock and Stephen Palmer; *Scientific American*, December 1990]. What Hebb did not notice, although the results were included in a paper he published at the time, was that the dark-reared rats took much longer than normal rats to learn to distin-

guish vertical from horizontal lines. Only many years later, after he had again changed his ideas about the relative importance of innate and learned mechanisms, did he appreciate the significance of this result.

Hebb received his Ph.D. from Harvard in the middle of the Depression, when there were no jobs in physiological psychology to be had. He therefore stayed on for a year as a teaching assistant, a post that enabled him to continue his work with Lashley. In 1937 there was still no improvement in the job market, but Hebb's luck held out. His sister was taking her Ph.D. in physiology at McGill and heard that Wilder Penfield, a surgeon who had just established the Montreal Neurological Institute there, was looking for someone to study the consequences of brain surgery on the behavior of patients. She passed on the information to her brother, and his application for the two-year fellowship was successful. He married again and returned to Montreal. The young man who thought he could run away from his family destiny and become a novelist found himself one of a medical group pioneering the treatment of neurological disorders.

Penfield's specialty was the treatment of focal epilepsy by surgically removing scarred areas of the cerebral cortex. He was acutely aware that he was operating on the organ of the mind and that a false move could deprive his patient of speech, intelligent behavior or even consciousness. Although Penfield was not a psychologist, his work exposed him to the relation between the mind and the nervous system. This experience no doubt influenced his decision to appoint psychologists to his team and explained the close interest he took in their findings.

Hebb's main responsibility was to study the nature and extent of any intellectual changes in patients consequent to cortical excisions. Such research was not new: it began after World War I with the psychometric testing of soldiers who had suffered penetrating head wounds and continued later in patients with brain tumors. In many cases, the lesions produced significant intellectual loss, but their locus and extent were difficult to determine. In contrast, surgical removals are more precisely defined, and epileptic scars do not cause the widespread damage that bullets or tumors do.

Hebb soon faced a peculiar problem. Psychologists then regarded the frontal lobes of the cerebral cortex as the seat of human intelligence, on the grounds that this region is relatively much larger than the corresponding areas in less intelligent animals. Yet Hebb was not able to detect intellectual loss in patients whose frontal lobes had been destroyed by accident or surgical necessity. This seeming lack of effect impressed Hebb deeply and inspired his quest for a theory of the brain and intelligent behavior.

Although his observations set him off on fruitful lines of inquiry, later work showed that Hebb had relied too heavily on standard intelligence tests. Brenda Milner, one of his students, who continued the work he had begun on Penfield's patients, found that frontal-lobe lesions often make it difficult for the patient to relinquish a behavior that has ceased to be appropriate. Although they may not be detected by intelligence tests, personality changes after frontal-lobe damage can profoundly affect the patient's life.

At the end of his fellowship at the neurological institute, Hebb finally found a permanent job at Queen's University in Kingston, Ontario. There, despite his

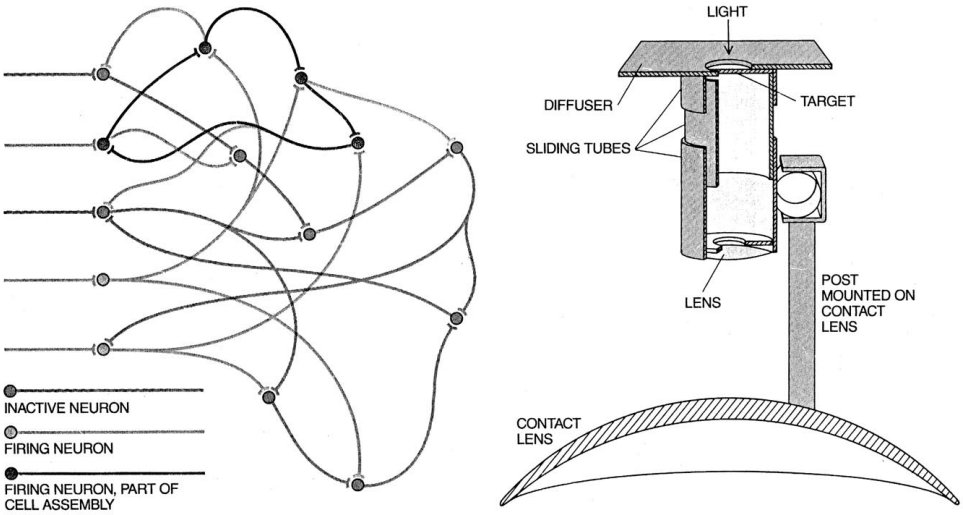


Figure 38.1
Hypothetical cell assembly begins with parallel fibers connecting input from the retina to corresponding points in the primary visual cortex. These neurons, in turn, connect to the "association" cortex. Converging input fires cells and activates closed loops. Synaptic changes ensue that enable the loop to fire with little input, producing output that represents to the brain what the eye has seen. Retinal fatigue supports the cell-assembly theory by causing images to fade in a peculiar fashion. The apparatus fixes an image on receptors until their signal decays. Then lines drop out, one or two at a time, until the figure is gone. Hebb argued that each line was represented by a neuronal feedback loop. When the retinal signal falls below the critical value, the loop stops oscillating, and the line disappears.

heavy teaching load, he kept up work on the problem of intelligence. Together with a student, Kenneth Williams, he developed a variable-path rat maze as an analogue to human intelligence tests. The Hebb-Williams maze was widely used for the next quarter century. But Hebb was proudest of a theoretical paper in which he proposed that adult intelligence was crucially influenced by experience during infancy, basing his argument on the results of his research at the Montreal Neurological Institute. The paper was virtually ignored at the time, although it is now accepted almost as a commonplace, having been embodied in such preschool enrichment programs as Head Start. But the concept was too advanced for its time: in 1940 most psychologists practically defined intelligence as an innate characteristic.

To reconcile his studies of childhood influences with the apparent harmlessness of frontal-lobe lesions, Hebb hypothesized that the region's main function was not to think but rather to facilitate the tremendous acquisition of knowledge during the first few years of life. Experiments to determine the relative effects of early and late brain lesions did not always support this idea, but it provided a stepping-stone to Hebb's later theories.

In 1942 Lashley became the director of the Yerkes Laboratories of Primate Biology in Florida, and he invited Hebb to join his research team to study chimpanzee behavior. Hebb jumped at the chance of doing full-time research with Lashley again, although he was not at first very enthusiastic about working

with chimpanzees. Lashley's intention was to develop tests of learning and problem solving for the animals, while Hebb would study their personalities and emotional characteristics. Then they would start a program to determine how brain lesions affected a range of variables.

The chimpanzees proved more difficult to train than Lashley had imagined. The delays meant that no brain operations were carried out during Hebb's tenure at Yerkes. Nevertheless, he was fascinated by his observations of chimpanzees and said he learned more about human personality in his five years of watching chimpanzees than at any other time since his own first five years of life. The apes manifested distinct personalities and a sense of fun that tended toward slapstick. Hebb and the other members of the staff derived a more cerebral amusement from the verbal contortions of orthodox behaviorist visitors as they attempted to describe the animals' practical jokes and broad clowning without resorting to "mentalistic" language.

Hebb's long and close observation of the many chimpanzees in the primate laboratory taught him that experience was not the only factor in the development of personality, including pathological manifestations such as phobias. He showed, for example, that young chimpanzees, born in the laboratory and known never to have seen a snake before, are frightened the first time they are shown one. Chimpanzees are also frightened of models of chimpanzee or human heads or other isolated body parts or of familiar caretakers wearing unusual clothing. Moreover, Hebb was one of the first to observe the social behavior of captive porpoises and to suggest that it implied a level of intelligence comparable to that of the apes. His observations may have influenced his later conclusion that level of play provides a good index of intelligence.

Lashley's interest in the ways the brain categorizes perceptions into knowledge about the world rekindled Hebb's curiosity about concepts and thinking. The problem can be rephrased as a question: How does the brain learn to lump one triangle, car or dog with another even though no two triangles, cars or dogs produce the same pattern of stimulation on sensory receptors?

The turning point came when Hebb read about the work of Rafael Lorente de Nó, a neurophysiologist at the Rockefeller Institute for Medical Research, who had discovered neural loops, or feedback paths, in the brain. Up to that point, all psychological theories, whether physiological or not, assumed that information passed through the organism along a one-way track, like food through the digestive system. Hebb recognized that Lorente's looping paths were just what he needed to develop a more realistic theory of the mind.

Feedback was not entirely new in learning theory. Almost all models assumed that the output of the organism influences the input in some way, for instance, by enabling the animal to receive a reinforcing stimulus. Unfortunately, feedback proceeding in this way, through a single path, would operate slowly and often unreliably. But with millions of internally connected feedback paths, it would clearly be possible to establish internal models of the environment that might predict the effects of possible responses without having to move a muscle.

Hebb's specialization in vision led him to concentrate his early neural theories on that system. Knowing that the point-to-point projection from the retina to

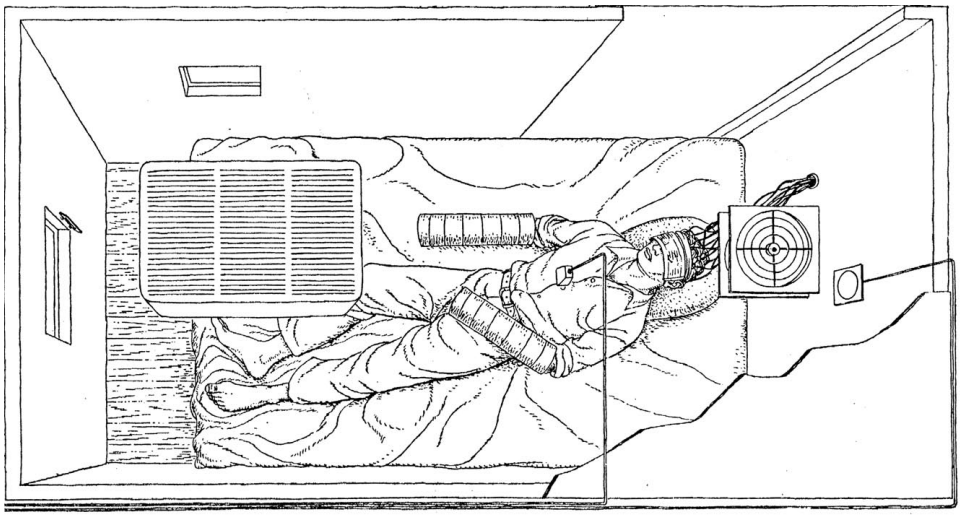


Figure 38.2
Isolation experiment carried the study of sensory deprivation beyond the realm of individual cell assemblies. Cuffs prevented touch, a plastic shield disrupted pattern vision and a U-shaped foam cushion attenuated sounds not masked by the air conditioner in the ceiling. EEG electrodes recorded the subject's brain waves and a microphone enabled him to report his experiences. The volunteers' ability to think deteriorated, and some of them even started to hallucinate.

the cortex does not extend beyond the primary visual cortex, he assumed that the neural relays projected into the surrounding cortex in random directions, thus scrambling the retinal pattern [see "The Visual Image in Mind and Brain," by Semir Zeki; *Scientific American*, September 1992]. Such an arrangement could recombine signals from different parts of the image—that is, they could converge on the same target neuron, causing it to fire. The resulting impulses could then return to the earlier neurons in the path, closing the feedback loops.

Repeated activation of any given loop might then strengthen that loop in the following way. If the axon of an "input" neuron is near enough to excite a target neuron, and if it persistently takes part in firing the target neuron, some growth process takes place in one or both cells to increase the efficiency of the input neuron's stimulation. Synapses that behave according to this postulate became known as Hebb synapses—somewhat to Hebb's amusement, it may be said, because this postulate is one of the few aspects of the theory he did not consider completely original. Something like it had been proposed by many psychologists, including Freud in his early years as a neurobiologist.

Nevertheless, Hebb's postulate was the most clear and formal statement, although in 1949 it was pure speculation. Since then, however, studies of single neurons have confirmed that synaptic strengths do change in some neurons in accordance with the postulate. Hebb may also have been correct about the mechanism of permanent change. A former student of his, Aryeh Routtenberg of Northwestern University, has recently pointed out that a protein associated with neuronal growth is produced when neurons are stimulated in ways that increase synaptic strength.

Hebb assumed that most of the synapses in the cortical lattice are initially too weak to fire spontaneously. To fire, they would require the converging of stimulation from a number of active neurons. Some neurons in the lattice receive converging inputs and thus fire when a particular pattern of neurons in the sensory cortex is fired by a stimulus. Some of the activated neurons have synaptic connections with one another, which are also strengthened whenever the stimulus is presented. Eventually the connections between the simultaneously firing neurons in the lattice become strong enough for them to continue firing one another in the absence of input from the stimulus, creating an internal representation of the stimulus, called a "cell assembly" by Hebb.

The concept of the cell assembly, in my view, was Hebb's greatest contribution to psychological theory, not to mention philosophy. It revived the 19th-century psychologists' attempt to explain behavior in terms of the association of ideas, a project that the behaviorists had derailed by arguing that "ideas" were no more real than the notion of little men inside the head. By so arguing, the behaviorists maintained that ideas, and thus mentalism, had no place in scientific psychology.

Unfortunately, few seemed to notice that the behaviorists replaced ideas with equally insubstantial constructs with misleading names, such as "stimuli" and "responses." These were not real events or chains of events but attributes that became associated with one another in some imaginary black box that scientists were forbidden to refer to as the brain. Hebb put a stop to this charade by showing, in principle at least, that ideas could have just as firm a physical basis as muscle movements. They could consist of learned patterns of neuronal firing in the brain, initially driven by sensory input but eventually acquiring autonomous status.

In its original form the neural theory was undoubtedly too simple to have worked. A major problem was that the cell assembly did not incorporate inhibition, because contemporary science did not recognize it. Sir John C. Eccles, a very influential neurophysiologist at the Australian National University in Canberra, was still vigorously denying the existence of inhibitory synapses. Moreover, many important connections of the neocortex had not yet been discovered, and the functional significance of the diversity of cortical neurons was only hinted at.

Without inhibiting factors, however, learning would strengthen synaptic connections until all neurons fired continuously, making the system useless. This effect was observed in computer models of the cell assembly, called conceptors, constructed in the 1950s by Nathaniel Rochester and his colleagues at the IBM research laboratory in Poughkeepsie, X.Y. Hebb himself seems never to have set finger to a computer to test his idea that random nerve nets could organize themselves to store and retrieve information. But such so-called neural nets later inspired many computer models, from the perceptron to parallel distributed processing, and have even found applications in industry.

By the time *The Organization of Behavior* reached publication, Hebb was back in Montreal as chairman of McGill's psychology department. Ten years later, when he stepped down as chairman, he had forged one of the strongest departments in North America. He found it easier to build what he wanted because the department was almost nonexistent when he began, and he turned out to be

adept at campus politics and soon discovered how to use his growing reputation to apply pressure where it would do the most good. It is perhaps significant that he was also one of the best chess players at the university.

Most of Hebb's research at McGill was related to his cell-assembly theory. Experiments to obtain direct physiological evidence for the theory were far beyond the scope of contemporary methodology. (They still are.) Instead he tested behavioral predictions of the theory. He tried, for instance, to strengthen his earlier conclusions on the influence of rearing on adult intelligence. Most of the results supported his theory that animals raised in an enriched, or more complex, environment would, in later life, outperform animals raised in bare cages.

There was one embarrassing exception. Litters of pure-bred Scotties were split, and half the pups were reared as pets in the homes of members of the staff and half were reared in cages in the laboratory. Hebb was not fortunate in the choice of his puppy, Henry. It was congenitally incapable of finding its way around, invariably got lost as soon as it was out of sight of the house and had to be recovered from the dog pound on several occasions. Naturally, Henry turned out to be near the bottom of the class when, as a full-grown dog, it was tested in a maze.

In a related series of experiments, Hebb investigated the effect of impoverished sensory input on the behavior of adults, including human volunteers [see "The Pathology of Boredom," by Woodburn Heron; *Scientific American*, January 1957]. Students were paid generously to undergo severe sensory deprivation for as long as they could stand it (none lasted even a week). Their ability to think began to deteriorate, and some of them even started to hallucinate. The Korean War was then in progress, and many workers attempted to use such isolation experiments to understand and combat the "brainwashing" techniques employed by the Chinese.

Hebb also pursued his old idea that early brain injury should be more damaging than injury in an adult. But the results were rendered uncertain by several factors, the most important being the capacity of the young brain to reorganize itself. For example, if an infant sustains an injury in an area of the left hemisphere that is important for speech in the adult, the right hemisphere takes over this function, and speech is not seriously impaired. But if an adult sustains damage in the same area, the result may be a permanent loss of language skills.

Because of such problems with the study of cognition, Hebb came to believe that the best evidence for the cell assembly came from experiments on retinal fading. Images of simple figures were projected onto the eye by a very small lens system attached to a contact lens, ensuring that the image always fell on the same place. As the receptor cells become fatigued, the image fades and disappears, but not all at once. Usually entire lines disappear suddenly, one or two at a time, until the entire figure is gone. Hebb explained the phenomenon by saying that each line is represented by neuronal activity circulating in a closed loop. The activity, once started, continues even after the input from the retina has decayed to a low value because of feedback around the loop. But at some critical value the reverberation stops abruptly, and the line disappears. These experiments do not provide conclusive evidence for the cell assembly as

Hebb envisaged it. Yet even if Hebb's version should turn out to be incorrect, it would not diminish the value of his idea that some neural activity continues to symbolize an object even after the object has stopped stimulating the sense organs.

Had *The Organization of Behavior* consisted only of the chapters in which Hebb criticizes current approaches and elaborates his cell-assembly theory, it is likely that few people would have read it. The book's appeal lies in its second half, in which Hebb discusses emotion, motivation, mental illness and the intelligence of humans and other species in the light of his theory. These essays are refreshingly forthright. On mental health, for example, Hebb wrote: "We still need an Ajax to stand up and defy the lightning and ask, What is the evidence? when some authority informs the public that believing in Santa Claus is bad for children, that comic books lead to psychological degeneracy, that asthma is due to a hidden mental illness."

Hebb built his department and his field by capturing the interest and imagination of the best students at an early stage. He taught the introductory course himself, making it immensely popular—at one point it numbered 1,500 students, about half the yearly undergraduate enrollment. Many future professors of psychology found their calling in these lectures. Like most of what Hebb did, his course was unique; no textbook at the time came close to including the material and ideas he dealt with, so he wrote his own. The first edition of *A Textbook of Psychology* appeared in 1958. In contrast to the majority of introductory texts of the day, it had more ideas than pictures.

Hebb also gave a graduate seminar that was attended by every psychology graduate student at McGill over a period of 30 years. It was famous not only for its stimulating discourse but also for Hebb's ever-present stopwatch and the slips of paper on which he noted incorrect pronunciations and other errors of presentation. It was Hebb's ambition never to have a McGill student overrun his or her allotted time at a meeting, and on the whole he was successful. McGill honored Hebb in 1970 by naming him chancellor; he became the only faculty member ever appointed to that position.

In 1977 Hebb retired to his birthplace in Nova Scotia, where he completed his last book, *Essay on Mind*. He was appointed an honorary professor of psychology at his alma mater, Dalhousie, and regularly participated in colloquia there until his death, at 81, in 1985.

Further Reading

The Organization of Behavior: A Neuropsychological Theory. D. O. Hebb. John Wiley, 1949.
Essay on Mind. D. O. Hebb. Lawrence Erlbaum Associates, 1980.
Parallel Learning in Brains and Machines. G. Ferry in *New Scientist*, Vol. 109, No. 1499, pages 36–38; March 13, 1986.
Textbook of Psychology. Fourth edition. D. O. Hebb. Lawrence Erlbaum Associates, 1987.
Mind and Brain. Special Issue of *Scientific American*, Vol. 267, No. 3; September 1992.

Chapter 39

Imaging the Future

Michael I. Posner and Daniel J. Levitin

One thousand years ago it was not universally held that the mind was located within the brain. One hundred years ago, the firm conviction that brain and mind were related led phrenologists to map the topography of the scalp and face (figure 39.1). In the last 10 years, cognitive psychologists studying mental operations have embraced neuroimaging techniques to localize mental operations in the brain, and to study their orchestration as humans perform a variety of tasks (figure 39.2). What will we find as scientists explore and chart the brain in the next 10 years, 100 years, or 1000 years?

Extrapolating the Current Scene

Before speculating about the future, it seems appropriate to begin with a brief account of what we already know (or at least the two of us think we know) of the brain through current methods. As we reach the last half decade of the 20th century it still amazes us that we can see pictures of our own minds at work. If a thought process can be sustained for only a few seconds, the snapshot revealed positron emission tomography (PET) or functional magnetic resonance imaging (fMRI) can show us which parts of our brain anatomy are active and to what degree. We know already that there are specific brain anatomies for reading (Posner & Raichle, 1994), listening to music (Marin, 1982; Sergent, 1993), mentally practicing your tennis serve (Roland, 1994), calculating numbers (Dehaene, 1995), and imagining a friend's face (Kosslyn, 1994). The methods for revealing the macroanatomy (in the range of millimeters to centimeters) of any mental process are clearly available.

Anatomy

One clear finding that emerges from these methods is that every cognitive task entails a particular network of brain areas; often we can link these brain areas to a specific computation required by the task. Some brain areas are very specific to a given cognitive domain so that they are only active if the task involves language or recognizing a face. Other brain areas appear to carry out very general computations that may be important in any task domain. For example, the lateral cerebellum appears active in both sensory and motor timing, as if it represented a central clock (Ivry & Keele, 1989).

From chapter 6 in *Mind and Brain Sciences in the 21st Century*, ed. R. L. Solso (Cambridge, MA: MIT Press, 1997), 91–102. Reprinted with permission.

Figure 39.1
A picture of classic phrenology. The areas of the brain come from studies of bumps on the head and the cognition represents the faculty psychology common at the turn of the century. (From Krech, D., and Crutchfield, R., *Elements of Psychology*. © 1958 by David Krech and Richard S. Crutchfield. © 1969, 1974 by Alfred A. Knopf, Inc.)

In the coming decades, we can expect our maps of brain anatomy to yield greater detail and spatial resolution, even if no new methods are invented. However, the attraction of young physicists formerly working on military problems to the study of the brain is such that we can be fairly certain that new and unanticipated ways of imaging brain activity will arrive. How should we use this increased resolution? Not just to make finer and finer maps! Rather, we need to seek principles of how important cognitive activity becomes distributed in brain regions.

In neuroscience, the cortical column is seen as the basic unit of organization of the human brain. Imaging methods have already shown that adjacent brain areas seem to become active as tasks change slightly. This forms the starting point for a principled approach to cortical organization at a macro level. The parietal lobe is involved in shifts of covert attention, but high up in its most superior regions it is active when the shift is to a purely visual event for which

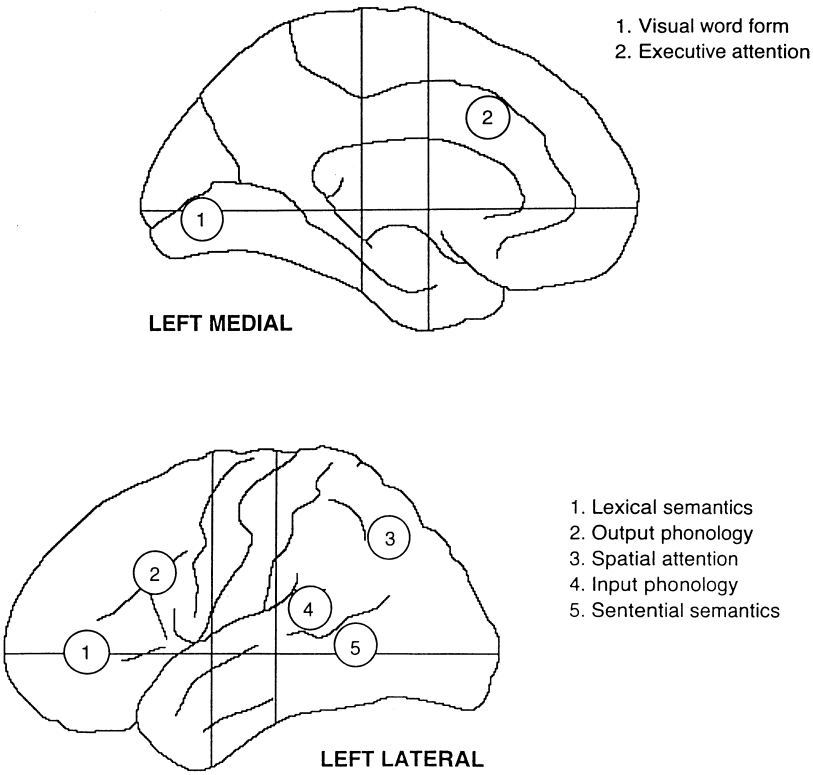


Figure 39.2
Modern phrenology. The areas of the brain summarize studies using PET and fMRI to observe changes in blood flow under experimental and control conditions. The cognition represents ideas of the types of computation involved in many cognitive tasks.

little response is required. When the shift of attention involves more detailed analysis of the nature of the visual event, or when overt orienting is allowed, the areas of activation appear to involve more inferior areas of the parietal lobe. When the task is purely one of recognition and no shift of attention is needed, the mid- to inferior temporal lobe is active.

Similarly, when a task is mostly passive (listening to a voice or music) different areas of the frontal midline show activity than those active when one begins to respond rapidly to a task (such as shadowing a word). Moreover, practicing a task can alter the brain areas involved. In one type of experiment, subjects are shown a noun and asked to respond with a word describing how it is used (e.g., "pound" to the word "hammer"). During this task, there are strong activations in the anterior cingulate, and in the left lateral posterior and anterior cortex. These activations disappear with practice but this is accompanied by an increase in activation in other brain areas that appear to be involved in automated (or "overlearned") tasks such as reading a word (Posner & Raichle, 1994). Perhaps our colleagues in the 21st century will be able to integrate these findings into a set of principles that will describe the organization of the brain for cognition.

Comparative Anatomy

As principles emerge in the study of the human brain these areas of activation can be viewed in relation to the known areas of primate brains to advance evolutionary analysis of cortical development. Some advances in this area have already taken place. For example, the visual word form area appears to involve portions of the brain that are also important for processing color. This relatively recent evolution suggests that the processing of visual words takes advantage of the high spatial frequency analysis available with the parvocellular areas of the visual system. Similarly, there is reason to think that the grammar of human language may take advantage of brain areas originally developed for hierarchical mechanisms of motor programming (Greenfield, 1991).

There are other advantages of exploring relationships between neuroimaging in human and animal models. The neuroimaging methods have been confined to an anatomy in the millimeter range, while cortical columns are in the micron range. The *micro*anatomy is important for understanding how *computations* are made by neurons. We already have some idea of the power of this method from studies of how mental rotations are computed in the monkey motor system (Georgopolis et al., 1989) and how perceptual motions are computed within area MT of the monkey (Newsome et al., 1994). The coming period should confirm and expand our knowledge of these mechanisms. We can then build on what we know about the details of neuronal computation in animals as additional constraints in the development of models of complex human tasks. As such models emerge it will be important to be able to examine the circuitry involved in human cognition to confirm predictions from models and to shape the agenda for the kinds of animal studies that will be needed.

Circuitry

Today it is also possible to observe the orchestration of many brain areas in real time. So far this has been accomplished mainly by relating the distribution of activity visible in event-related electrical and magnetic fields to generators found active in anatomical studies (Snyder et al., 1995).

Circuitry and Reading One of the areas for which the most knowledge is already available is in reading of words. During the reading of a foveal word (Posner et al., 1996), computations occur in the right posterior occipital lobe at about 80 ms that relate to *features* of the word. By 130 to 180 ms the "visual word form" of the left posterior cortex is activated. For simple, clearly visible words, this is followed by activity in a frontal midline attention system by 170 ms and in a left lateralized frontal semantic system by 220 ms.

These activations contribute to organizing the saccade for the next fixation which typically begins by about 270 ms (Posner et al., 1996). It is known from cognitive studies that the saccade is influenced by knowledge of the meaning of the current word. Therefore, it is necessary that information about the meaning of the word be available before the eyes are moved. Before the saccade begins there is more activation in anterior semantic areas related to word meaning, as well as higher-level frontal attentional areas. We should not think that these activations are purely in the direction of posterior to frontal. Rather there is feedback of information from frontal systems into posterior areas. Thus, in

imagining a scene, frontally based attention and semantic systems can be used to activated posterior areas related to the visual form of the scene (Kosslyn, 1994).

While we can expect some of the details of these findings will likely be modified by future work, the way is clearly open for a detailed description of the time courses of mental operations in high-level human tasks.

There is also much scope for improvement in our ability to image non-invasively the circuitry involved in brain activity. While at present, electro-and magnetic activity recorded from outside the head are all that are available, the development of new statistical tools, including Bayesian analyses to constrain the solution space, will allow us to project the probable three-dimensional source of activation deep into the brain (Tucker et al., 1994). Future advances in dense sensory array measurement (e.g., 128 or 256 channels) of the brain's electrical and magnetic fields at the head surface promise new insights into the sources of these fields.

By combining the surface measurement with accurate information on the tissues of the head and brain from magnetic resonance imaging (MRI), the electrical (electroencephalogram or EEG) and magnetic (magnetoencephalogram or MEG) studies may be guided by additional constraints for localizing the neural sources. Unlike metabolic or blood flow (PET or fMRI) methods that have a poor temporal resolution, the EEG and MEG techniques provide a millisecond temporal resolution that is better suited to the time course of cognitive operations performed by neural circuits.

Plasticity

In cognitive science there has been a long-standing interest in the nature of expert performance (Chi et al., 1988). These studies show that there are major differences in representation of the same information by novices and by experts, and changes in representation that accompany the development of expertise within an individual. Very familiar to most cognitive psychologists is the impressive achievements of chess masters. Simon has estimated that this skill is based on many thousands of hours of practice and produces an elaborated semantic memory that allows reproduction of the chessboard in lawful master-level games (Chase & Simon, 1973). Chase and Ericsson (1982) have observed these changes in memory with practice in students trained to have numerical digit spans of up to 100 items. In these cases we do not yet know how the brain is altered by the experience involved.

While it has not yet been possible to understand the achievements of chess masters in *neural* terms, some studies of the neural basis of expert performance have already taken place. Familiar to most cognitive psychologists is the phenomenological change that accompanies learning to read. The ability of a skilled reader to recognize each letter of a lawful word at a lower threshold than the letter in isolation shows that learning the skill provides a visual chunk that eliminates the need to scan and integrate the letters. For this effect we already have a candidate neural system in the left medial occipital lobe (Posner & Raichle, 1994) that appears to be involved in performing this recognition function. It appears that this skill requires years of practice and produces signs of an adult "word form system" only at about age 10 (Posner et al., 1996).

The studies outlined above suggest the continued plasticity of some aspects of brain circuitry with new learning. However, there is already evidence of critical periods in the learning of skills. Weber-Fox and Neville (1996) studied the learning of English by immigrants from China who came to the United States at ages ranging from 2 years to adulthood. They found that the brain circuitry involved in understanding the meaning of lexical items was similar regardless of age of immigration. However, the circuitry underlying grammatical judgments resembled American natives for those who immigrated as young children, but was very different in those whose immigration was late. A similar critical period has now been reported in learning the violin. Children who begin lessons prior to age 12 show changes in somatosensory cortical representation between the left and right hands that are not present even in expert violinists who began their lessons late (Elbert et al., 1995).

At present we have only a rudimentary understanding of how the anatomy, circuitry, and plasticity of the brain are involved in the performance of high-level human skills. It is clear that the accuracy and replicability of these findings is likely to improve steadily as new methods and more laboratories examine the results. However, it appears unlikely that we will ever be able to describe playing chess, for example, in terms of every brain area of computation that is invoked during a masters game. What will our goals be then and what progress toward their attainment can we expect?

Dynamic Brains

The study of psychology during the period from World War II to the mid-1980s was a study of how information was transferred between people and within a person. Psychology then was the study of the logic of how information was perceived, transformed, stored, and communicated. The brain was a black box, opaque to the physical substrate required to perform the functions specified by psychological models of mental events. A dominant metaphor was that psychologists studied software and for the logic of the programs it really didn't matter what hardware was required to run them. The current scene that we have described above—in which the hardware is also of interest—was ushered in by two related events. First, methods of neuroimaging opened up the human brain to investigation. It was now possible to image parts of the brain and see how they cooperated during performance. Second, a new class of models were developed, based on the idea of complex computations resulting from simple neuronlike units. These two events have allowed psychologists to describe the anatomy, circuitry, and plasticity of higher forms of human performance. In this section we try to speculate on what the consequences of this new opportunity will be.

A series of very important studies by Merzenich and colleagues (Merzenich & Sameshima, 1993) has found that the brain of the sensory systems of higher primates can change with experience. What is new as the century draws to a close is our capacity to also observe these changes in humans as they acquire skills.

We have barely begun to understand the capacity for change in the human brain. In a recent functional magnetic resonance study (Spitzer et al., 1995) showed evidence that brain areas that coded the concept "animal" were sepa-

rate from the brain areas that were responsive to pictures of furniture. Whereas the areas active for these concepts were generally located within brain areas related to semantic processing, the number, the exact location, and the extent of the activation appeared to differ among people. Putting these observations together with the learning-dependent changes in brain maps shown in Merzenich's work, we may expect that spending a month furnishing your apartment would lead to an expansion of "furniture" representational areas in your brain, while working in a zoo might change the extent and depth to which animals are represented in the brain. These findings might well explain the common observation that our thoughts and even our dreams tend to be dominated by events related to current experiences—observations that on a more micro scale are seen in laboratory studies of priming.

Learning

Cognitive science, which views humans as intelligent, learning, and thinking creatures, is beginning to have an influence in the field of education. To bridge the gap between theory and practice in this important arena, a number of cognitive psychologists have moved into the classroom. A recent book (Bruer, 1993) describing the significance of cognitive work for classrooms has received an award from the American Federation of Teachers.

We believe that in the future the field of cognitive neuroscience will be likely to also have a large impact on education. This may seem at first a somewhat unrealistic idea. There have been so many false starts, so many pop theories of brain functions, that many people (perhaps even the two of us) are wondering if we can learn things about the brain of sufficient importance to describe to those entrusted with the education of children. Nonetheless, we think that the new methods available to us both in terms of cognitive theory and brain imaging are stronger then ever before and we really must attempt to relate our findings to educational issues.

Recovery of Function

Possibly the first area to benefit from the study of brain imaging will be the field of cognitive retraining following strokes or other closed head injuries. There has been evidence of some success in attempting to improve outcome from new forms of learning. However, since the mechanisms of recovery are not known, it has proved difficult to know whether these improvements in behavior are related to the training or due to spontaneous recovery that may also occur with delay after the injury. The ability to image the brain should allow much more detailed evidence of what the learning might do to change the anatomy or circuitry involved in cognitive tasks. In time we should know whether—and under what conditions—the relearning influences recovery within the damaged tissue, allows new areas to take over, or produces wholly new strategies that involve very different brain areas than those involved in the original task.

School Subjects

Already some tasks involving reading, music, and arithmetic have been studied in terms of anatomy and circuitry. Is there anything likely to emerge in cognitive neuroscience that will influence how these subjects are taught? One recent

report illustrates what might be possible. Dehaene (1996) has argued that areas of the posterior parietal cortex are important for understanding the *quantity* of a number. He argues that this area of the brain is active when subjects are required to compare quantity, and moreover, lesions of this area produce a deficit in comparing and otherwise understanding quantity. Dehaene argues that this area may be common to both humans and animals and underlies our ability to know about quantity.

Griffin et al. (1994) have argued that children who are at risk of failing arithmetic in elementary school have a deficit in understanding the quantity of numbers so that they are unable to compare numbers. When this deficit is corrected by intensive education, they show marked improvement in their ability in arithmetic courses. These findings raise the possibility that we may be able to detect difficulties in comprehension related to specific brain areas and perhaps observe changes in activation of these areas that occur following the training. If so, our ability to diagnose a wide variety of learning disabilities in children may improve and benefit from neuroimaging in much the same way as described above for recovery of function following brain damage.

Individuality

The science of human differences has been heavily influenced by psychometric methods on the one hand, and on the other by the promise of twin studies that have suggested the genetic basis of personality. Work at three different levels of understanding in particular holds great promise: (1) genetic approaches, including the human genome project, (2) neuroimaging, and (3) phenotypic approaches to defining personality. As these methods are refined and the different levels related to one another, there is the promise of new excitement in the study of individual differences in cognition, emotion, and personality.

Genetic Level

According to recent estimates, the full sequence of the human genome will be completed ahead of schedule, by 2005. We now know that the brain has 3195 distinctive genes, and that roughly 17% of these are involved with cell signaling. It is conceivable that in the near future we will have found connections between particular genes in the brain and individual differences in personality traits. Whether particular genes will indicate a propensity for certain behaviors or determine those behaviors will undoubtedly be the subject of much popular debate. However, the currently available evidence—based on studies of identical twins separated at birth—is quite convincing that genetics is not deterministic of behavior; it merely provides a statistical model that accounts for only a portion of behavior variability (Lykken et al., 1992; Lykken et al., 1993), and then only for the behavior of groups, not individuals. Thus, although certain gene markers might become associated with the potential for particular behaviors, the existence of a particular gene will not likely determine one's behavior.

What we still do not know much about is the way in which genes are translated first into biological substrates in the brain, and then into psychological mechanisms, such as a trait, nature, attitude, or preference. Moreover, we still know very little about the relation between traits and behavior, as the power of

situational forces can often confound our predictions based on traits (Malle, 1995; Ross & Nisbett, 1991). The findings of behavior geneticists and personality and social psychologists will need to be integrated in the coming years to advance our understanding of these issues.

Neuroimaging

The genome findings, taken in concert with imaging studies, promise to illuminate the anatomical basis for many types of individual differences. The development of fMRI allows ready superposition of changes in blood flow and brain structure. Thus we can see how activation of brain areas relates to the structure of individual brains. We have already reviewed evidence that the structure and function of the brains of violinists differ if practice is started early enough (Elbert et al., 1995). We should be able to determine which differences depend upon practice and which may involve genetic differences that perhaps lead to the acquisition of high-level skill. In current cognitive psychology both genetic and learning views of individual differences have advocates; it seems likely that the use of imaging methods will provide a basis for separating and relating these approaches.

Phenotypic Structure

Although we use thousands of words to describe how people differ from one another, mathematical analyses show that our perception of human traits clusters in an orderly fashion, such that most of the traits on which people differ can be described by a location in a five-dimensional coordinate system, the "Big Five" personality model (Goldberg, 1993). This finding seems to hold up across a variety of cultures and languages, adding to the growing body of evidence that the strong version of the Whorf-Sapir hypothesis is untenable.

A subset of work on personality differences concerns one particular constellation of traits, those associated with what we loosely call "intellect." The recent, more inclusionary definitions of "intelligence" that allow for athletic, spatial, artistic, and other "nonacademic" intelligences (Gardner, 1983) broaden our notions of what it means to be intelligent. These new definitions also provide an expanded framework for the study of expertise. The near future may see changes in how we teach our children, as a result of the formal acknowledgment by academia that disparate forms of accomplishment exist.

Sociopathy

An example of how these three levels of research are merging comes from recent studies on criminal and aggressive behaviors. Geneticists have speculated that an "aggression" or "criminality" gene may soon be found. fMRI studies of the brains of murderers have shown clear differences in blood flow between them and normals: murderers tend to show far less frontal lobe activity, a possible indicator that they are less able to regulate feelings of aggression in a normal way. Obviously this evidence is merely correlational, and it does not demonstrate a causal link. Yet, some researchers believe that violent behavior will turn out to be physiologically determined. Raine (1993) predicts that the next generation of clinicians and the public will "reconceptualize non-trivial recidivistic crime as a psychological disorder."

At the phenotypic level, the constellation of traits that seem correlated with criminality appear clustered along the negative axis of one of the Big Five dimensions conscientiousness/undependability. The degree to which criminal behavior is a matter of genetics, anatomy, environment, or personality is a problem that may become subject to scientific resolution. A recent, forward-looking integration of many of these ideas in sociopathy may be found in Lykken (1995).

Some have predicted that within 10 years we will be able to actually diagnose those people with a propensity for committing violent acts before they have committed them, possibly during childhood or preadolescence (Gibbs, 1995). How this information is to be used will undoubtedly become a source of considerable public debate in the coming decades, and psychologists will likely be called upon to participate in this debate. But any "individual differences screening" based on anatomical or genetic markers can yield only statistical probabilities for a group. That is, we might be able to say that *X*% of a group that shows the propensity for violence will go on to commit violent acts, but we cannot predict with any certainty *how a given individual will behave*. Consequently, the most responsible use of such information might be never to gather it in the first place. It is our worst fear that screening information might be used to force medical interventions or incarceration on individuals who have demonstrated only that they are part of a group with a statistical chance of violent behavior, a course that would parallel the ugly history of the eugenics movement in the United States in the early 1930s. A concomitant fear is that future public policy might ignore the findings of science: even seemingly benign interventions that result from the best intuition and intentions can backfire (McCord, 1978).

The one thread common to these three approaches to the study of individuality seems to be an emerging consensus that the brain contains a great deal of "hard-wiring" of systems that are specialized for particular functions, or the expression of particular behaviors. But this hard-wiring is only a framework, one that holds tremendous plasticity, and is malleable as a result of experience and environmental input. Although the range of human differences appears infinite, these differences are contained within a system that is finite in its genetic, anatomical, and phenotypic description.

Theory of Consciousness

The coming decades should hold more interaction among researchers in the various fields that study human behavior. The neuroimaging methods have already brought together many fields in an effort to map the human brain. One theoretical topic that has united philosophy with the sciences is the effort to understand the physical basis of our conscious experience.

The question of what it is to be conscious has recently again become a central one in many serious scientific circles. Proposals range from the anatomical—for example, locating consciousness in the thalamus or in thalamic-cortical interactions—to the physical—for example, the proposal that consciousness must rest on quantum principles. Will all of these speculations provide a basis for

understanding the centuries-old philosophical problems of how our mental experiences arise and how they relate to the brain?

One aspect of experience that has traditionally been related to or equated with consciousness is attention (James, 1890/1950). The images of human brains at work have revealed brain areas that seem closely related to programming the order of our mental computations. The areas responsible for programming amplify particular computations or suppress others, and they comprise various networks supporting selective attention. So far, these studies have supported three fundamental working hypotheses that together constitute current efforts to produce a combined cognitive neuroscience of attention. First, the brain possesses an attentional system that is anatomically separate from the various data-processing systems that can also be activated passively by visual, auditory, and other input. Second, attention is accomplished through a network of anatomical areas; it is neither the property of a single brain area nor is it a collective function of the brain working as a whole. Third, the brain areas involved in attention do not carry out the same function, but specific computations are assigned to specific areas (Posner & Raichle, 1994).

One major source of our feelings of conscious control involves the act (or illusion!) of voluntary control over behavior and thought. Volitional control is by no means total as the (presumably unwanted) tendency of depressed people to dwell on negative life events clearly shows. Yet all normal people have a strong subjective feeling of intentional or voluntary control of their behavior. Asking people about goals or intention is probably the single most predictive indicator of their behavior during problem solving. The importance of intention and goals is illustrated by observations of patients with frontal lesions (Duncan, 1994) or mental disorders (Frith, 1992) that cause disruption in either their central control over behavior or the subjective feelings of such control. Despite these indices of central control, it has not been easy to specify exactly the functions or mechanisms of central control.

Nonetheless there are some cognitive models of executive control that outline subsystems serving to control cognitive processing (Norman & Shallice, 1986). According to this model, attentional systems involve two qualitatively different mechanisms. The first level of control corresponds to routine selection (contention scheduling) in which the temporarily strong activity wins out. However, when a situation is novel or highly competitive (i.e., requires executive control), another supervisory system would intervene and provide additional inhibition or activation to the appropriate schema for the situation. Norman and Shallice (1980, 1986) have argued that the supervisory system would be necessary for five types of behaviors or situations in which the routine or automatic processes of the contention scheduling mechanisms would be inadequate and executive control would be required. These are (1) situations involving planning or decision making; (2) situations involving error correction; (3) situations where the response is novel and not well-learned; (4) situations judged to be difficult or dangerous; and (5) situations that require overcoming habitual responses.

One of the most interesting findings from the era of neuroimaging is that tasks involving these properties have all activated areas on the midline of the frontal lobe (Posner & DiGirolamo, 1998). Moreover, lesions in this general area

produce a remarkable loss of spontaneous thought and action. Damasio (1994) has recently described the effects of lesions of this area as follows: "Their condition is described best as suspended animation, mental and external—the extreme variety of an impairment of reasoning and emotional expression. Key regions affected by the damage include the anterior cingulate cortex, the supplementary motor area, and the third motor area." While more recent studies of surgical lesions of this area have not produced the devastating loss of mental function, so we do not know the extent or the neural system involved.

A new debate has emerged over whether consciousness is a function or a process, and thus over whether consciousness will be found to exist in a particular place in the brain. Elsewhere, one of us has argued that the anterior cingulate is likely to be a necessary and important component of tasks that are associated with consciousness (Posner, 1994), but that consciousness is a distributed, multifaceted function. The other of us has argued the not inconsistent idea that consciousness is an emergent property of the brain-as-a-whole, and that it is a *process*, not a *thing* (Luu et al., 2001). Thus, just as we don't expect to find "gravity" at a particular location in the middle of the earth, we shouldn't expect to find consciousness at a particular place in the head.

We can only speculate about the consequences of these new developments in the theory of attention for philosophical views about the relationship of brain to mental experience. Although we feel some confidence about the scientific predictions made in this chapter, we have relatively little idea what effect they might have upon the philosophical disputes that have attended the issue of consciousness. However, we can express our hope that the new developments in neuroimaging that will take place over the coming decades might help psychologists and philosophers to overcome the inhibitions of the hundreds of years of separation between mental and physical events. With an understanding that knowledge of the brain's anatomy provides constraints for more conceptual—or traditional cognitive—models, the psychologist and the philosopher will thus be able to reason, each from his or her understanding of neuroscience and of cognition. This joint approach will provide the basis for understanding the mechanisms of awareness and cognitive control as elements of consciousness.

References

Bruer, J. T. (1993). *Schools for thought: A science of learning in the classroom*. Cambridge, MA: MIT Press.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16). New York: Academic Press.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: G. P. Putnam.

Dehaene, S. (1996). The organization of brain activations in number comparisons: Event related potentials and the additive-factors method. *Journal of Cognitive Neuroscience*, *8*, 47–68.

Duncan, J. (1994). Attention, intelligence and the frontal lobes. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–734). Cambridge, MA: MIT Press.

Elbert, T., Pantex, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science, 270*, 305–306.

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Hillsdale NJ: Erlbaum.
Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
Georgopoulos, A. P., Lurito, J. T., Petrides, M., & Schwartz, A. B. (1989). Mental rotation of the neuronal population vector. *Science, 243*(4888), 234–236.
Gibbs, W. W. (1995). Seeking the criminal element. *Scientific American, 272*(3), 76–83.
Goldberg, L. R. (1993). The structure of phenotypic personality traits. Presented at Sixth European Conference on Personality. *American Psychologist, 48*(1), 26–34.
Greenfield, P. M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences, 14*(4), 531–595.
Griffin, S., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice*. Cambridge, MA: MIT Press/Bradford Books.
Ivry, R. B., & Keele, S. W. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience, 1*(2), 136–152.
James, W. (1890/1950). *The principles of psychology*. New York: Dover.
Kosslyn, S. (1994). *Image and brain*. Cambridge, MA: MIT Press.
Luu, P., Levitin, D. J., & Kelley, J. M. (2001). Brain evolution and the process of consciousness. In P. G. Grossenbacher (Ed.), *Consciousness and brain circuitry: Neuro-cognitive systems which mediate subjective experience*. Philadelphia: John Benjamins.
Lykken, D. T. (1995). *The antisocial personalities*. Hillsdale, NJ: Erlbaum.
Lykken, D. T., Bouchard, T. J., McGue, M., & Tellegen, A. (1993). Heritability of interests: A twin study. *Journal of Applied Psychology, 78*(4), 649–661.
Lykken, D. T., McGue, M., Tellegen, A., & Bouchard, T. J. (1992). Emergenesis: Genetic traits that may not run in families. *American Psychologist, 47*(12), 1565–1577.
Malle, B. F. (1995). The person and the situation: Conceptual issues in theories of social behavior. Unpublished manuscript.
Marin, O. S. M. (1982). Neurological aspects of music perception and performance. In D. Deutsch (Ed.), *The psychology of music*. San Diego: Academic Press.
McCord, J. (1978). A thirty-year follow up of treatment effects. *American Psychologist, 33*(3), 284–289.
Merzenich, M. M., & Sameshima, K. (1993). Cortical plasticity and memory. *Current Opinion in Neurobiology, 3*(2), 187–196.
Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
Newsome, W. T., Shadlen, M. N., Zohary, E., Britten, K. H., & Movshon, J. A. (1994). Visual motion: Linking neuronal activity to psychophysics performance. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 401–413). Cambridge, MA: MIT Press.
Norman, D. A., & Shallice, T. (1980). *Attention to action: Willed and automatic control of behavior* (Technical Report No. 99). Center for Human Information Processing.
Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1–18). New York: Plenum Press.
Posner, M. I. (1994). Attention: The mechanism of consciousness *Proceedings of the National Academy of Sciences of the United States of America, 91*(16), 7398–7402.
Posner, M. I., Abdullaev, Y. G., McCandliss, B. D., & Sereno, S. E. (1996). Anatomy, circuitry, and plasticity of reading. In J. Everatt (Ed.), *Visual and attentional processes in reading and dyslexia*. New York: Routledge.
Posner, M. I., & Raichle, M. E. (1994). *Images of mind*. New York: Scientific American Library.
Posner, M. I. & DiGirolamo, G. J. (1998). Executive attention: Conflict, target detection and cognitive control. In R. Parasuraman (Ed.), *The attentive brain*. Cambridge: MIT Press.
Raine, A. (1993). *The psychopathology of crime*. New York: Academic Press.
Roland, P. (1994). *Brain activation*. New York: Wiley-Liss.
Ross, L., & Nisbett, R. E. (1991). The person and the situation: Perspectives of social psychology. New York: McGraw-Hill.
Sergent, J. (1993). Mapping the musician brain. *Human Brain Mapping, 1*, 20–38.
Shallice, T. (1988). *From neuropsychology to mental structure*. New York: Cambridge University Press.
Snyder, A. Z., Abdullaev, Y., Posner, M. I., & Raichle, M. E. (1995). Scalp electrical potentials reflect regional cerebral blood flow responses during processing of written words. *Proceedings of the National Academy of Sciences of the United States of America, 92*, 1689–1693.

Spitzer, M., Kwong, K. K., Kennedy, W., Rosen, B. R., & Belliveau, J. W. (1995). Category-specific brain activation in fMRI during picture naming. *NeuroReport, 6*, 2109–2112.

Tucker, D. M., Liotti, M., Potts, G. F., Russell, G. S., & Posner, M. I. (1994). Spatio-temporal analysis of brain electrical fields. *Human Brain Mapping, 1*, 134–152.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8*, 231–256.

Index

Absent-minded errors, 391
Absolute pitch, 300, 303–304, 306, 490–491, 522
Abstraction, 277, 279, 285
Adaptation, 665–679
Aerial perspective, 505
Affordance, 423–425
Agnosia, auditory, 303
Akinetic mutism, 827
Alexia, 302, 303, 825
Alzheimer's disease, 769, 787
Ambiguity. *See also* Illusion
 perceptual, 139–142, 182
 of stimuli, 139
American Psychological Association (APA), 125
American Sign Language (ASL), 262, 689, 696
Ames room, 173
Amusia, 300, 302, 764, 792
ANOVA, 127
Aphasia, 302, 805
Apparent motion, 165, 225, 229–231
Apraxia, 765
Artificial categories, 261
Artificial intelligence (AI), 37–38, 44–46, 50–51, 53, 57, 95–111, 266, 399, 567, 831
Assimilation principle, 328–329, 350
Atkinson, R. C., and R. M. Shiffrin, 296
Attention, 148–158, 363–395, 404, 405, 413, 843, 851
 for action, 823–824, 827
 action-slips, 389–393
 automatic processing, 384–389
 bottleneck theories of, 150, 381–382
 disengagement of, 376
 divided, 157, 369, 378
 dual-task performance, 378–384
 engagement of, 376–377
 evaluation of theories of, 393–394
 filter theory of, 150
 focused, 408
 focused auditory, 363–369
 focused visual, 370–378
 goal-directed, 148
 perceptual, 148–158, 185
 preattentive processing, 153–155, 158
 selective, 148, 152, 829
 shifting, 376–377
 visual/spatial, 821
Attentional beam, 370
Attentional engagement theory, 373, 374–375
Attributes, perceived, 264
Auditory event, 219
Auditory scene analysis, 213–248
Auditory streams, 220, 225, 229
Autoassociation, 468–471
Automatic process, 387, 388, 393
Automaticity, 388–389
Automatons, 12
Avocational pursuits, 772
Awareness, 3, 148, 153, 171, 184, 296, 445, 447

Backpropagation, 473
Balint's syndrome, 277
Base-rate frequency, 586
Basic level objects, 251, 253, 254–260, 265, 267, 268
Basic object research, 256
Basilar membrane, 218
Bayes' rule, 127, 586
Bayesian inferencing, 127, 586
Behavioral efficiency, 392–393
Behaviorism, 7, 8, 16, 831
Belief, 338, 339, 341
Belongingness, principle of, 222, 224
Berkeley, Bishop George, 6
Best-examples model, 280–281, 285, 287, 289–290
Bestist, 758
Between-group trials, 199
Between-subjects design, 123, 124, 125
Biases, 585–600
 effectiveness of a search set, 592
 imaginability, 593
 retrievability of instances, 592
Big five factor model of personality, 849–850
Bilingual, 693
Binet, Alfred, 781–782
Binocular cues, 166–168
Binocular disparity, 166–167
Binocular parallax, 505
Black box, 846

Block-recognition problem, 231
Bottleneck theories. *See* Attention
Bottom-up processing, 140, 176–177, 185, 408, 510
Brain simulator reply (Berkeley and M.I.T.), 103–104
Broadbent, D. E., 150, 365–367
Broca's area, 767
By-products, in evolution, 644–649, 673

Carolina Abecedarian Project, 798
Carryover effect, 124
Cartesian mental substance, 110
Categorization, 251–269, 277–287, 290–291
 advance information, 262
 critical assumption theory of, 278
 horizontal dimension, principle of, 259–261
 implications of principles for other fields, 258–259
 problematic issues of principles, 264–266
 vertical dimension, principle of, 254–258
Category
 subordinate, 258, 268
 superordinate, 254, 255, 267, 268, 280, 283
Category resemblance, 254, 255, 260
Causal theories, 19
Cell assembly theory, 831, 837, 839
Central capacity theory, 382–383
Characteristic frequency, 455
Chase-Simon theory, 540
Chinese room, 95–111
Chi-square test, 127
Choice, rational theory of, 601
Chunking, 297, 526, 537–540, 568
Circuitry, brain, 844, 847
Clever Hans, 121–122
Clinical judgement, 622–623
Closure, 158, 194, 232–233, 402–403, 407
Coarse coding, 456
Codes, internal, 820
Cognitive economy, 252, 259
Combination reply (Berkeley and Stanford), 104–105
Combinatorial explosion, 42–43
Common attributes, 255–256
Common fate principle, 163, 193, 197
Common region principle, 195–196, 197, 200, 201
Communication, 446
Compatibility effects, 611–612
Competitive learning rule, 84
Complexes, 538
Computer, 35–58, 75, 107, 109–111, 122, 444–445, 447, 450, 452
 modeling, 237
 simulations, 526
Concave function of money, 602
Conceptors, 837
Conceptual integration, 655
Conceptual models, 426–434
Conceptually driven processing, 140, 176–177, 408
Concomitants, 675
Conditioning, 832
Confidence, 339–341
Confidence intervals, 127, 128
Confidentiality, 126
Conflict, 614–617
Conjunctive events, 595
Connection strength (between units in PDP models), 80. *See also* Parallel distributed processing (PDP)
Connection weights (between units in PDP models), 83. *See also* Parallel distributed processing (PDP)
Connectionist models, 62–66, 68, 72, 75, 79–80, 86–89, 314, 455, 465, 569, 831, 837
Consciousness, 10, 11, 16, 51, 296, 363, 850–852
 criteria for behavioral similarity, 11–12
 criteria for visual similarity, 12
Constancy. *See* Lightness constancy; Loudness constancy; Orientation constancy; Shape constancy; Size constancy
Constraints, 466
 adaptation, 646
 continuity, 70
 human information processing, 526
 multiple simultaneous, 58, 60
 multiple simultaneous in motor control, 65
 mutual simultaneous in word recognition, 60–61
 optimal design, 645–647
 semantic, 301
 stability, 518–519
 uniqueness, 70
Constructionism, 311–352
Contention scheduling, 388
Context
 effects, in categorization, 289–291
 influence of, in perception, 180–184
 model, in categorization processes, 284, 287
 role of, in basic-level objects and prototypes, 265
Continuation, law of good, 194, 206
Continuity principle, 70
Contour
 melodic, 299–300
 subjective, 160, 511
Controlled experiment, 118, 120
Controlled search process, 385
Controlled studies, 116–119
Convergence, 166
Conversational implicature, 725–732. *See also* Grice, H. P.; Maxims; Obscurity of statement

Cooperative principle, 724–730
Co-opted adaptation, 648
Co-opted spandrel, 648
Correlation, 128, 780, 795–797
Correlational studies, 119–120
Correlational theories, 19
Cortical lattice, 837
Creole, 696–698
Critical period hypothesis of language learning, 694–696
Critical period hypothesis of skill learning, 846
Crystallizing experiences, 773
Cued recall tests, 315
Cue validity, 254, 255, 260, 269
CYRUS, 45

Darwin, Charles, 640–641, 645
Data-driven processing, 140, 176–177, 185, 408
Debriefing, 125
Decision making, 448, 585–600
Deep structure, in auditory scene analysis, 238
Delta rule, 84, 469
Demand characteristics, 124
Dennet's "good city" test, 40, 41
Dependent variable, 117
Depth cues, 165, 170
Depth perception, 166, 207
 mechanisms, 68
Descartes, René, 37, 110, 685
Description invariance, 607
Descriptive studies, 120–121
Design
 experimental, 115–130 (*see also* Between-subjects design; Within-subjects design)
 principles of (in product design), 419
Detection task, 374
Determinism, 626
Development, cognitive (effect on categorization), 258, 261
Diary studies, 389–390
Dichotic listening, 150–151, 365
Differential motion cue, 171
Differential reproductive success, 641
Digit-span task, 543
Discrimination failures, 389
Dishabituation, 244
Disjunctive events, 595
Disorders of visual attention, 376–377
Disparity (binocular), 166–167
Distal stimulus, 138–139, 143
Distance cue, 170
Distinctive events, 322–323, 325–326
Distinctive features, 254, 404
Droodles, 176
Dualism, 3–7, 110–111. *See also* Mind-body problem

Early vision, 404–405
Ebbinghaus, H., 314–315, 322
Echoes, 241
Ecological optics theory, 147
Ecological validity, 223, 347
Effect sizes, 127
Eidetic imagers, 327
Elaborative rehearsal, 329
Electroencephalogram (EEG), 845
Element aggregation, 197
Element connectedness, 195–197, 200
Eliciting preference, 610
Emotion, 495–496
Empiricism, 143
Encoding failure, 296
Encoding specificity, 341, 350
Environment of evolutionary adaptedness (EEA), 643–644
Epiphenomenalism, 5
Eugenics, 850
Events
 disjunctive, 595
 distinctive, 322–323, 325–326
 recent, 322
 role of objects in, 266–268
Evidence-processing system, 234
Evolution, 239, 639–660, 665–679
Exaptation, 639–660
Exclusive allocation, principle of, 221
Executive routines, 316
Expectations, 180–184
Experience error, 191
Experimental conditions, 118
Experimental design, 115–130
 design flaws, 121–122
 ethical considerations, 125–126
 types, 123–125
Expert systems, 44

Failure, 623–624
Family of meanings, 272–275
Family resemblances, in categories, 262, 272, 280–281
Fan effect, 346, 348–349
Feature detector, 455
Feature integration theory, 372–374
Feature maps, 404
Feedback paths, neural loops, 835
Feedback (principle of, in HCI), 437
Figural goodness, 160–161
Figurative language, 733–747
Figure/ground, 158–160
Filling in, 70, 73, 232–235, 510–511
Fitness (classical), 641
Fitness theory (inclusive), 641
Fixation, 163
Focal epilepsy, 833
Focal instances, 280

Foils, 326
Forces of attraction, 228
Foresight, 621
Forgetting, 296, 298–299, 311–312, 314, 324, 343–349, 350
Formal modeling, 622–623
Frame, 61
Framing effect, 605–609
Free recall tests, 315
Frequency detector, 455–457
Frequency proximity, 236
Frontal lobes, 769–770, 833
F-test, 127
Function, biological, 650–651
Functionalism, 9, 110
Functional magnetic resonance imaging (fMRI), 841, 845, 848–849

G (general) factor, 784
Gambler's fallacy, 589, 628
Ganglion retinal cells, 142
Gedanken experiment, 96
Generalizability, 118
Geons, 178
Gestalt
 principles of grouping, 162, 193–200, 203, 206, 222, 224, 227–228, 230–233, 247, 512 (*see also* Proximal stimulus; Proximity model)
 psychology, 146, 214, 237, 242, 399
Gibson's ecological optics, 147
Gist, 298–300, 321
Good continuation, 194, 206
Gradient of size, 505
Grasping, 58
Grice, H. P., 723–732. *See also* Conversational implicature; Maxims; Obscurity of statement
Grouping-by-attraction, 234
Guided search, 154

Habituation, 244
Hallucinations, 142
Harmony, 237
Hebb, Donald O., 83–84, 87, 831–839
 rule, 84, 87
 synapse, 831, 836
Hebb-Williams maze, 834
Hedges, 262
Hemholtz, Hermann von, 145, 504, 512
Hereditarian theory, 779, 794, 797–799, 808
Hermann grid, 142
Heuristic, 237, 585–600
 adjustment and anchoring, 594–595
 availability, 592
 old-plus-new, 223
 representativenes, 585–591
 rules, 585
Higher-level cognitive processes, 136
Horizontal dimension of category systems, 253
Human genome, 848
Human subjects, 125
Hyperbole, 728
Hypnosis, 337
Hypothesis-driven processing, 140, 176–177, 408

Idealism, 6
Idiomaticity, 747
Idioms, 733–747
Idiot savants, 536, 570, 572, 763
Illusion, 139, 142–143, 225, 401. *See also* Ambiguity
 of continuity, 225, 234, 246
 duck/rabbit, 140–141
 Ebbinghaus, 144
 geometrical optical, 142
 impossible figures, 164
 misassignment, 225
 Müller-Lyer, 144
 Necker cube, 140–141
 perceptual, 139
 Poggendorf, 144
 Ponzo, 170–171
 top hat, 144
 of validity, 590
 vase/faces, 140–141
 Zöllner, 144
Illusory conjunctions, 156–158, 373, 400, 402, 408, 410
Illusory correlation, 594
Imagery, 258
Imitation game, 35
Implicit knowledge, 464
Impossible figures, 164
Independent groups design, 123
Independent variable, 117
Indeterminacy, 634
Induced motion, 163
Infant perception of musical structure, 122
Inference, 145
Information-processing theory, 553, 667, 784, 788, 819
Informed consent, 125
Inheritance, 640
Inhibition
 between-level, 71
 competitive, 71
Inner ear, 218
Inspection time task, 785–786
Intellect, 849
Intelligence, 448, 520–522, 833. *See also* IQ test; Scholastic Aptitude Test (SAT); Wechsler adults intelligence scales (WAIS); Wechsler intelligence scales for children (WISC)
 assessment, 774–775

Cattell Culture Fair test of, 789
G (general) factor, 784
genetic basis of, 806–807
multiple intelligences theory, 762, 772, 776
neurophysiological basis of, 787–788
raw, 766
unitary, 783–789, 794
Intentionality, 4
Interactionism, 5
Interference, 344–349, 350, 380, 463
Interposition, 169, 171
Invariance, 350
Invariant pitch-class representation, 458
Invariants of experiences (regarding memory), 320, 321
Inverted spectrum argument, 12–13
IQ tests, 520–522, 753, 766, 780–809. *See also* Intelligence
Isomorph, 559–563
Item output, 261–262

Johnston and Heinz theory, 368
Just-so stories, 652

Knowledge, 295–296, 551–558
Körte's third law, 229–231

Language, 243, 258, 271, 685–702
development, 690–702
logic of natural language use, 262–264
Lateral inhibition, 142
Law of small numbers, 589
Learning, 80, 84, 243, 466, 528–546, 634, 835, 847
explicit rule formation, 80
Hebbian, 466–467
language, 685–702
sequences, 471
speed of, 261
Level of abstraction, 251, 254, 255, 266, 267, 268
Levels of processing, 329–330
Lightness constancy, 202
Linear perspective, 169, 171, 505
Linguistic input, 691, 693. *See also* Language
Linguistic structure, 691. *See also* Language
Logic, 262
Loss aversion, 601–610
Loudness, 299, 307
Loudness constancy, 506
Luminance, 202–203, 208
LUNAR, 41
Lyrics, 300–301, 307

Magnetoencephalogram (MEG), 845
Maintenance rehearsal, 329
Many mansions reply (Berkeley), 106–111
Mapping, 419, 426, 433, 434, 435
natural, 436
Masking, 234
Matching, 410, 820
Materialism, 6
Maxims (Gricean) 723–732. *See also* Conversational implicature; Grice, H. P.; Obscurity of statement
Measurement error, 126
Meiosis, 728
Melody, 230, 237, 299–300, 304, 306–307
Memory, 75–80, 153, 158, 182, 295–297, 311, 448–452, 463, 471, 525–546
accuracy, 339–341
autobiographical, 325, 577
episodic/explicit, 311
eyewitness, 334, 341
filter theory of, 366
flashbulb, 324
long-term, 296, 303–306, 388, 536
melody, 301
muscle, 306
music, 299
photographic, 326–328
recollection, 311–312, 316, 320, 324, 326, 332–333, 337, 341, 344–345, 350, 367
reconstruction, 316, 321–322, 328, 333–335, 338–339, 341, 347, 351
record-keeping theories of, 311–352
retrieval, 388–389
short-term (working memory), 537, 784, 786, 827
source, 335–336
systems, 295
working, 537, 784, 786, 827
Mental set, 182, 363
Mental states
epistemic status of, 9
ontological status of, 9
Mentalism, 831
Metaphor, 733–747
Meta-analyses, 127
Methodism, 633
Method of loci, 326
Metonymy, 745–746
Microanatomy, 844
Mind-body problem, 3. *See also* Dualism
Mind-brain identity theory, 6
Mnemonic technique, 326–327
Mode of control, 391
Models
neural net (of memory), 75–80
parallel distributed processing (*see* Parallel distributed processing (PDP))
pattern associator, 81–87
perception, 75
Modular theory, 383–384
Monism, 6
Motherese, 687

Motion cues, 166–168
Motion parallax, 168, 505
Motion perception, 164–165
Motivation, 395
Motor systems, 65, 182, 256, 765
Multiple intelligences theory, 762, 772, 776
Multiple realizability, 9
MUSACT model, 462, 468, 475
Mutation, 643, 669
MYCIN, 45

Nativism, 143, 146
Natural selection, 503, 639–673
Nature and nurture, 143
Neural loops, 835
Neural networks, 62–66, 68, 72, 75, 79–80, 86–89, 314, 455, 465, 569, 831, 837
Neural relays, 836
Neural theory, 837
Neuroimaging, 787, 824, 841–852
Noise, 645
Norman and Shallice theory, 387–388
Null hypothesis, 127

Object recognition, 178–180, 189–210, 399–413
Obscurity of statement (Gricean principle), 730. *See also* Conversational implicature; Grice; Maxims
Occlusion, 169, 171, 233
Octave equivalence, 483
Old-plus-new heuristic, 223
Ontogeny, evolutionary adaptation, 642
Ontological problem, 3–12
Operationalism, 40
Optimal design, 645–647
Order effects (in experimental design), 124
Organism design theory, 667
Orientation, 163
Orientation constancy, 174–176
Oscillatory circuits, 459
Other minds, problem of, 11
Other minds reply (Yale), 105
Overlap principle, 341–344, 350

p value, 127
Pain, phantom limb, 5
Parallel distributed processing (PDP), 62–66, 68, 72, 75, 79–80, 86–89, 314, 455, 465, 569, 831, 837
 activation of units, 71
 active representation, 79
Parallelism, 194
Parallel search, 153
PARRY, 43
Partition, 210
Past experience, 205
Pattern associator models, 81–87
Patterning ability (raw), 772
Pavlovian conditioning, 832
Peak experiences, 577
Perceived (world) attributes, 264
Percept, 158
Perception, 68, 70, 71, 133–188, 215, 258
 analytic stage of perception, 146
 color, 14
 depth, 166, 207
 direct, 505
 Gestalt, 468
 identification, 136–148, 176–186
 model, 75
 music, 302, 455–477, 481–498, 503–514
 object recognition, 136–148, 153, 176–186
 perceptual organization, 135–148, 158, 185–186, 242, 244
 of pitch, 460
 sensation, 135–148, 185
 shape, 160–161
 stereoscopic depth, 68, 89
 visual, 213
Perceptron, 469
Perceptual attention, 148–158, 185
Perceptual completion, 70, 73, 232–235, 510–511
Perceptual constancies, 171–172
Perceptual continuity, 410
Perceptual decomposition, 235
Perceptual grouping principles, 161–163, 191–207
Perceptual organization, 135–148, 185, 189–191
 objects and scenes, 189– 210
 region analysis, 207–210
Perceptual set, 182
Perceptual stream, 246
Perceptual unity, 410
Peripheral vision, 405
Personality, 769, 848
 big five factor model of, 849–850
Phantom limb pain, 5
Phasic response, 459
Phenomenological criteria, 15
Phi phenomenon, 165
Phoneme restoration effect, 70, 177
Phonological processing, 826–827
Photo bias, 334
Phrenology, 842–843
Physiologizing, 831
Pick's disease, 769
Pictorial cues, 168–171
Pidgin, 696–698
Pilot study, 125
Pitch, musical, 299–300, 303–306, 481–498
 absolute, 300, 303–304, 306, 490–491, 522
Plasticity, 845
Polysynthetic language, 710
Pop out, 402, 404–405
Population, statistical, 123

Positron emission tomography (PET), 787, 824, 841, 845
Pragnanz law, 163
Preattentive processing, 153–155, 158
Preference reversal, 611–612
Presentism, 633
Primal sketch, 399
Primary auditory cortex, 457, 826
Primary visual cortex, 824
Priming, 262, 372, 411, 462–465
Primitive segregation, 242
Principles of exclusive allocation, 221–225
Prior probability, 586–587
Privacy (in human subjects testing), 126
Proactive interference, 345
Probability distribution (subjective), 596
Problem solving, 343–344
 strategies, 558–563
Procedure invariance, 610
Prodigies, 570–572, 763
Productivity curve, 558
Programme assembly failures, 389
Progressive supranuclear palsy, 376
Prominance hypothesis, 612
Proofreader's error, 568
Property, emergent, 4
Propositions, general form, 271
Prosody, 303
Prosopagnosia, 302
Prospect theory, 603
PROSPECTOR, 45
Prototypes, 251, 253, 259, 262–269, 461
Proximal stimulus, 138–139, 143
Proximity, 193, 195, 196, 197, 198, 199, 203, 227
 model, 279
 stimuli, 200
 Wertheimer's law of, 161
Psychophysical complementarity, 242
Psychophysics, 213–214

Quick-probe assumption, 37, 39, 40, 47

Random assignment, 116–118
Random-dot stereogram, 68–70
Random effects, 645
Random sampling, 118
Random selection, 118
Rational theory of choice, 618
Raven's matrices test, 787–788
Reaching, 58
Reaction time, 261, 822
Receptive field, 142
Recognition, 324–325, 330, 344, 351, 410
Recognition network, 413
Recognition tests, 315
Reference frames, 160–161
Refractory period, 381
Region analysis, 207–208
Region segregation, 158–159
Relative size, 169, 171
Relearning paradigm, 315
Remembering, 298, 312, 316, 318, 326, 334
Repeatability, 128
Repeated measures design, 124
Repetition detection task, 200
Repetition discrimination task, 199, 200, 201
Repetitious events, 322
Representation
 distributed, 80–81
 local, 80–81
 mental, 503
Representative sample, 123
Retinal cues, 505
Retrieval failure, 297
Retroactive interference, 345
Reverberation, 510
Rhythm, 299, 301, 488, 494–495
Risk, 602
Robot reply (Yale), 102–103

"S" (S.V. Shereshevskii), 326
Sample size, 123
Scene analysis, 216, 230–231, 239–240, 246
Schema, 62, 464
Schema theory, 392
Schema-based integration, 245
Schema-based segregation, 242
Scholastic Aptitude Test (SAT), 753. *See also* Intelligence
Script, 61
Second language learning, 695
Second law of thermodynamics, 509
Segmentation cues, 463
Segregation, 239
Selection, 640
Selectivity of processing, 363
Self-statement, 774
Semantic processing, 371
Semantics, 58, 60, 107, 109, 330, 410
Sensory buffer, 296
Sensory transducers, 503
Sequential integration, 236
Sequential organization, 235
Serial search, 154–155, 404
Sex hormones, 804–806
Shadowing, 150–151, 367
Shape constancy, 172–174
Shape perception, 160–161
Shiffrin and Schneider theory, 385–387
Significance testing, 127
Sign-language, 262, 689, 696
Sign test, 128
Similarity, 163, 193, 195, 197, 200, 232, 257, 285, 290
Simultanagnosia, 376
Simultaneous integration, 236

Situation awareness, 445, 447
Size constancy, 172–174, 213, 506
Size/distance relation role, 171
Socialization, 576
Sociopathy, 849–850
Spatial code, 458
Spatial inversion, 507–510
Spatial processing, 767, 802, 805
Spatial summation, 465
Special design, 644
Spectral organization, 235
Spectrogram, 217
  neural, 218
Spoken-language, 689. *See also* Language
Spontaneous speech, 692
Stanford-Binet test, 782
Stereogram, random-dot, 68–70
Stereoscopic depth perception, 68–69
Stereoscopic vision, 68
Stimulus-driven capture, 148–149
Storage failures, 389
Stratified sample, 118
Stream segregation, 226
Streaming effect, 225, 235–236, 246
Stroop effect, 384
Structure-emotion, 565, 576, 579
Subroutine failures, 389
Summary information, 290–291
Summary representation, 279, 281, 289–291
Supervenience, 10–11
Supervisory attentional system, 388
Symbol system, 772
Symmetry principle, 194, 206
Synchrony principle, 195
Synesthesia, 326
Syntactic theory, 237
Syntax, 58, 60, 107, 109, 238
Synthesis theory, 384
Synthetic stage of perception, 146
Systems reply (Berkeley), 99–102

Tachistoscope, 224
Tacit knowledge, 791
Task analysis, 534
Taxonomy, 251
Tempo, 299, 303, 306–307
Temporal composites, 455–480
Temporal reversal, 508
Temporal summation, 465
Test battery, 755
Test failures, 389
Testist, 758
Texture gradients, 171
Thalamus, pulvinar nucleus, 377
Think-aloud protocol, 524
Timbre, 214, 236, 299–300, 303, 307
Timing, 122
Timing deformations, 575
Tonality, 455–480
Tonic response, 459
Top-down processing, 140, 176–177, 408, 510
Transfer appropriate processing, 330
Transpositional invariance, 475
True experiments, 116–119
T-test, 127
Tuning, abstract feature, 456
Tuning curve, 455–456
Turing, Alan, 35, 37–40, 42, 48–50
Turing machine, 35
Turing test, 16–17, 35–53, 100, 110

Unconscious inference, 146, 504–506
Uniform connectedness, 208–210
Unit formation, 197
Units. *See also* Parallel distributed processing
  hidden, 471

Value (subjective), 602–604
Variance, analysis of (ANOVA), 127
Variation, evolutionary, 640
Vector spaces, 459–460
Veridical expectancies, 471
Vertical dimension of category systems, 251, 253, 254–260, 265, 267–268
Vibrato, 514
Visual adaptation, 400
Visual completion, 203
Visual processing, 399, 411
Visual search, 372, 402
Visual system, 457
Vocational, 772
Von Restorff effect, 323

Wagenaar, W., 326
War of the Ghosts, 315–317, 328
Weber's law, 405
Wechsler adults intelligence scales (WAIS), 782
Wechsler intelligence scales for children (WISC), 782
Wertheimer, Max, 161
Westist, 758
Winograd, Terry, 38
Within-subjects design, 124, 125, 127
Wittgenstein's puzzle, 474
Word association network, 827
Word recognition, 60–61
World correlational structure, perception of, 252–254

Zoom-lens model, 370